Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering

Dissertation

Enumeration of Factors in Special Languages

Prague 2021 Josef Rukavička

Acknowledgments

I wish to thank Štěpán Starosta and Edita Pelantová for their help and guidance. Also I appreciate the friendly environment of Department of Mathematics of FNSPE of CTU (Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague).

I acknowledge support by the Czech Science Foundation, grant GAČR 13-03538S, and by the Grant Agency of the Czech Technical University in Prague, grants No. SGS14/205/OHK4/3T/14 and No. SGS20/183/OHK4/3T/14.

Bibliografický záznam

Autor	Ing. Josef Rukavička České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra matematiky
Název práce	Enumerace faktorů jazyků speciálních typů
Studijní program	Aplikace přírodních věd
Studijní obor	Matematické inženýrství
Školitel	doc. Ing. Štěpán Starosta, Ph.D. České vysoké učení technické v Praze Fakulta informačních technologií Katedra aplikované matematiky
Školitel specialista	prof. Ing. Edita Pelantová, CSc. České vysoké učení technické v Praze Fakulta jaderná a fyzikálně inženýrská Katedra matematiky
Akademický rok	2020/2021
Počet stran	164
Klíčová slova	kombinatorika na slovech, enumerace faktorů, palindromy, bohatá slova, mocnin-prostá slova, privilegovaná slova, uzavřená slova, de Bruijnovy grafy, disekce nekonečných jazyků, palindromická délka

Bibliographic entry

Author	Ing. Josef Rukavička Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Mathematics
Title of Dissertation	Enumeration of Factors in Special Languages
Degree Programme	Application of Natural Sciences
Field of Study	Mathematical Engineering
Supervisor	doc. Ing. Štěpán Starosta, Ph.D. Czech Technical University in Prague Faculty of Information Technology Department of Applied Mathematics
Supervisor specialist	prof. Ing. Edita Pelantová, CSc. Czech Technical University in Prague Faculty of Nuclear Sciences and Physical Engineering Department of Mathematics
Academic Year	2020/2021
Number of Pages	164
Keywords	combinatorics on words, enumeration of factors, palindromes, rich words, power-free words, privileged words, closed words, de Bruijn graphs, dissection of infinite languages, palindromic length

Abstrakt

Tato disertační práce se zaměřuje na jazyky bohatých, mocnin-prostých, uzavřených a privilegovaných slov.

Ukážeme, že počet bohatých slov má subexponenciální růst a že faktorová komplexita bohatých slov je shora omezena kvazipolynomiální funkcí $c_1 n^{c_2 \ln n}$, kde c_1, c_2 jsou nějaké konstanty. Dále odvodíme horní mez pro počet uzavřených a privilegovaných slov dané délky.

V roce 1985 publikovali autoři Restivo a Salemi pět otevřených problémů, které se týkaly mocnin-prostých slov. V této disertaci částečně vyřešíme dva z těchto problémů. Nechť u je doprava rozšiřitelné α-mocnin-prosté slovo a v je doleva rozšiřitelné α-mocnin-prosté slovo nad abecedou s q písmeny, kde $\alpha > 2$ a $q \geq 3$. Potom ukážeme, že existuje slovo w takové, že uwv je rovněž α-mocnin-prosté slovo nad stejnou abecedou.

Je známo, že pokud w je bohaté slovo, tak existuje písmeno a takové, že wa je rovněž bohaté slovo. Odvodíme několik netriviálních výsledků pro bohatá slova, která lze rozšířit na bohatá slova nejméně dvěma různými způsoby.

Pro zadaná bohatá slova u, v ukážeme algoritmus, který rozhodne, jestli existuje bohaté slovo w takové, že u, v jsou faktory w.

Představíme další tři výsledky, které přímo nesouvisí s těžištěm této disertace. Tyto výsledky se týkají palindromické délky, de Bruijnových grafů a disekce nekonečných jazyků.

Tato disertace je koncipována jako soubor devíti autorových původních článků doplněný integrujícím textem. Sedm z nich již bylo publikováno v recenzovaných časopisech a dva jsou v recenzním řízení.

Abstract

This dissertation focuses on languages of rich, power-free, closed, and privileged words.

We show that the number of rich words grows subexponentially and that the factor complexity of rich words is bounded by a quasi-polynomial function $c_1 n^{c_2 \ln n}$ for some constants c_1, c_2. We derive an upper bound for the number of closed and privileged words.

In 1985, Restivo and Salemi published five open problems concerning power-free words. We solve partially two of these problems. To be specific, we show that if u is a right extendable α-power-free word and v is a left extendable α-power-free word over an alphabet with q letters, where $\alpha > 2$ and $q \geq 3$, then there is a word w such that uwu is also α-power-free over the same alphabet.

It is known that if w is a rich word then there is a letter a such that wa is also rich. We prove some nontrivial results describing rich words that can be "richly" extended in at least two ways.

For given two rich words u, v, we show how to decide whether there is a rich word w such that w contains u, v as factors.

Three other results are presented that are not in the main focus of the dissertation. These results deal with a palindromic length, de Bruijn graphs, and a dissection of infinite languages.

The dissertation is formed as a collection of nine author's original articles accompanied with an integrating text. Seven of them have already been published in refereed journals and two articles are currently in review.

Contents

List of original publications

We present nine articles that we have written during the doctoral study. The articles are sorted in ascending order by the first submission date. To be able to easily distinguish our articles from other cited sources, double square brackets are used when referring to them.

[[Ru01]] J. Rukavicka: *Bijections in de Bruijn Graphs*, Ars Combinatoria, Volume CXLIII, January, 2019, pp. 215-226.

[[Ru02]] J. Rukavicka: *On the Number of Rich Words*, In: Charlier É., Leroy J., Rigo M. (eds) Developments in Language Theory. DLT 2017. Lecture Notes in Computer Science, vol 10396. Springer, Cham.

[[Ru03]] J. Rukavicka: *Upper Bound for Palindromic and Factor Complexity of Rich Words*, RAIRO-Theor. Inf. Appl., 55 (2021) 1.

[[Ru04]] J. Rukavicka: *Rich Words Containing Two Given Factors*, In: Mercaş R., Reidenbach D. (eds) Combinatorics on Words. WORDS 2019. Lecture Notes in Computer Science, vol 11682. Springer, Cham.

[[Ru05]] J. Rukavicka: *A Unique Extension of Rich Words*, `https://arxiv.org/abs/1910.00826`, Submitted 2019.

[[Ru06]] J. Rukavicka: *Upper bound for the number of closed and privileged words*, Information Processing Letters, Volume 156, 2020.

[[Ru07]] J. Rukavicka: *Transition Property for α-power-free Languages with $\alpha \geq 2$ and $k \geq 3$ Letters*, In: Jonoska N., Savchuk D. (eds) Developments in Language Theory. DLT 2020. Lecture Notes in Computer Science, vol 12086. Springer, Cham.

[[Ru08]] J. Rukavicka: *Palindromic Length of Words with Many Periodic Palindromes*, In: Jirásková G., Pighizzini G. (eds) Descriptional Complexity of Formal Systems. DCFS 2020. Lecture Notes in Computer Science, vol 12442. Springer, Cham.

[[Ru09]] J. Rukavicka: *Dissecting Power of a Finite Intersection of Context Free Languages*, `https://arxiv.org/abs/2006.15160`, Submitted 2020.

1 Introduction

The origin of combinatorics on words, as an area of discrete mathematics, is often associated with the study of square-free words by the Norwegian mathematician Axel Thue at the beginning of 20$^{\text{th}}$ century [7]. In 1906, Axel Thue proved the existence of infinite square-free words on an alphabet with three letters [43]. A word w is *square-free* if w does not contain a factor of the form vv, where v is a nonempty word. Since the topic was accepted by the mathematicians as quite natural and interesting, the article [43] gave an impulse to generalize the result for both an arbitrary alphabet and an arbitrary power v^{α} of the factor v, where $\alpha > 1$ is a rational number. Such words are called power-free or α-power-free. (The formal definition of a power-free word will be given in Section 2.)

Up to now, power-free words remain one of the major themes in the area of combinatorics on words. Nowadays, some other generalizations of power-free words are being studied; for example abelian powers, pseudo squares, and reverse powers [12, 26, 33]. We omit the definition of these generalizations.

Next to power-free words and all their generalizations, combinatorics on words includes many other themes [16, 24, 25]. A full list of the themes would be quite extensive. Instead, we restrict our attention to the topics that are relevant for this dissertation.

In general, a lot of open problems in combinatorics on words deal with enumeration of some languages. A language is a set of finite words. For a given language L, enumeration of the language L usually means to find the number of words from L of a given length. Instead of enumeration of the language L, we say also that we enumerate words from L. Formally, we look for the function

$$f(L,n) = |\{w \mid w \in L \text{ and } |w| = n\}|.$$

For many languages, enumeration is a hard problem. This is why we are often satisfied with some lower and upper bounds for $f(L,n)$.

Factor complexity of a finite or infinite word w is the function $f(L,n)$, where L is the set of all factors of w. In our dissertation we construct upper bounds for the function $f(L,n)$, where L is the language of rich, closed, and privileged words. Also we derive an upper bound for the factor complexity of rich words.

Another general problems of combinatorics on words are so-called transition property and extendability of words. Understanding transition property and extendability of words from a given language L helps also with enumeration of the language L. This is how transition property and extendability fit in this dissertation.

Given a language L and two words $u, v \in L$. Transition property deals with following problems:

- Is there a word w such that $uwv \in L$?
- Is there a word $\bar{w} \in L$ such that u, v are factors of $\bar{w}$?

- Construct the words w and $\bar{w}$ if they exist.

Simply said, we ask for existence of a "transition" word and for examples of transition words. Note that in case of the second problem, the occurrences of u and v in $\bar{w}$ may overlap with each other and the order of occurrences of u, v in $\bar{w}$ does not matter. In our dissertation we investigate the transition property of power-free and rich words.

In case of extendability, we are concerned with following two problems:

- Given a language L and a word w, are there words words u_1, u_2, u_3, u_4 such that $u_1w, wu_2, u_3wu_4 \in L$?
- Construct the words u_1, u_2, u_3, u_4 if they exist.

Analogously like in the case of transition property, one asks for both existence and examples of the "extensions". A factorial language is a language L such that if $w \in L$, then all factors of w are also in L. In case of the factorial language L, it is sufficient to consider that $w \in L$ and u_1, u_2, u_3, u_4 are letters. In our dissertation, we are mainly concerned with languages of rich and power-free words. Both these languages are factorial.

The three mentioned topics, enumeration, transition property, and extendability of rich and power-free words are the main topics of this dissertation. Several other topics are included. We address these remaining topics in dedicated sections.

The dissertation has the following structure. Section 2 defines the basic notions that the dissertation deal with. The remaining sections focus on a summary of the results of our articles, known results and the context of the topics being researched. The layout of sections, topics, and articles is structured as follows:

- Section 3: Enumeration
  - Enumeration of rich words [[Ru02]].
  - Factor complexity of rich words [[Ru03]].
  - Enumeration of closed and privileged words [[Ru06]].
- Section 4: Transition Property and Extendability
  - Transition property of power-free languages [[Ru07]].
  - Extendability of rich words [[Ru05]].
  - Transition property of rich words [[Ru04]].
- Section 5: Palindromic length [[Ru08]].
- Section 6: De Bruijn sequences and de Bruijn graphs [[Ru01]].
- Section 7: Dissection of infinite languages [[Ru09]].

In Section 8 we suggest several topics for future research based on our results.

2 Preliminaries

Let A_q be a finite alphabet with $q \geq 1$ letters. Given a positive integer n, let

$$\mathrm{A}_q^n = \{a_1a_2\dots a_n \mid a_i \in \mathrm{A}_q \text{ and } i \in \{1,2,\dots,n\}\}.$$

We define that $\mathrm{A}_q^0 = \{\epsilon\}$. The element ϵ is called *the empty word*. Let $\mathrm{A}_q^+ = \bigcup_{n\geq 1}\mathrm{A}_q^n$ and let let $\mathrm{A}_q^* = \mathrm{A}_q^+ \cup \{\epsilon\}$. The elements of A_q^n are called *words* of length n, where n is a nonnegative integer. The length of the word $w \in \mathrm{A}_q^*$ is denoted $|w|$. It is well known that $|\mathrm{A}_q^n| = q^n$, where n is a nonnegative integer.

Let A_q^∞ denote the set of all infinite words over the alphabet A_q; formally

$$\mathrm{A}_q^\infty = \{a_1a_2\cdots \mid a_i \in \mathrm{A}_q \text{ and } i \geq 1\}.$$

Given a finite word $u \in \mathrm{A}_q^*$ and a finite or infinite word $w \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$, let uw denote the concatenation of the words u and w. We have that $\epsilon u = u$, $u\epsilon = u$, and $\epsilon w = w$.

Given a finite word $u \in \mathrm{A}_q^*$ and a finite or infinite word $w \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$, we say that u is a *factor* of w if there are $w_1 \in \mathrm{A}_q^*$ and $w_2 \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$ such that $w = w_1uw_2$. We say that u is a *prefix* of w if there is $w_1 \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$ such that $w = uw_1$. We say that u is a *suffix* of $v \in \mathrm{A}_q^*$ if there is $v_1 \in \mathrm{A}_q^*$ such that $v = v_1u$.

Let F_w denote the set of all factors of a finite or infinite word $w \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$. It follows that $\epsilon \in \mathrm{F}_w$ and if w is finite, then also $w \in \mathrm{F}_w$. In addition, let $\mathrm{F}_w(n) = \mathrm{F}_w \cap \mathrm{A}_q^n$, where n is a nonnegative integer; $\mathrm{F}_w(n)$ is the set of all factors of length n of the word w.

Let $w = w_1w_2\cdots w_{n-1}w_n \in \mathrm{A}_q^n$, where $w_i \in \mathrm{A}_q$ and $i \in \{1,2,\dots,n\}$. We define that $w^R = w_nw_{n-1}\cdots w_2w_1$ and $\epsilon^R = \epsilon$. The word w^R is called *reversal* of the word w. We say that a set $S \subseteq \mathrm{A}_q^*$ of finite words is *closed under reversal* if $w \in S$ implies that $w^R \in S$. We define that a word $w \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$ is closed under reversal if the set F_w is closed under reversal. It is easy to see that if $\mathrm{F}_w(n)$ is closed under reversal then $\mathrm{F}_w(j)$ is closed under reversal for each $j \leq n$.

A word $w \in \mathrm{A}_q^*$ is called a *palindrome* if $w = w^R$. (It follows that the empty word ϵ is a palindrome.) If w is a finite or infinite word, u is a palindrome, and $u \in \mathrm{F}_w$, then we say that w contains a palindromic factor u. A finite word $w \in \mathrm{A}_q^n$ is called *rich* if w contains $n+1$ palindromic factors (including the empty word) . It is known that a word w of length n can contain at most $n+1$ palindromic factors [15], hence rich words are those that contain the maximal number of palindromic factors. An infinite word $v \in \mathrm{A}_q^\infty$ is rich if all factors of v are rich.

A nonempty word w is a *border* of the word u if $|w| < |u|$ and w is both a prefix and a suffix of u. A word u is *closed* if there is a border w of u such that u has exactly two occurrences of w. It follows that w occurs only as a prefix and as a suffix of u. A word u is *privileged* if $|u| \leq 1$ or if u contains a privileged border w that appears exactly twice in u. Obviously privileged words are a subset of closed words.

There is a connection between rich and privileged words. It was shown that every word w of length n contains $n+1$ distinct privileged factors. If the set of privileged factors of w coincides with the set of palindromic factors, then w is rich [30].

An α-*power*, where $\alpha \geq 1$ is a rational number, of a nonempty finite word r is the word $r^\alpha = rr \dots rt$ such that $\frac{|r^\alpha|}{|r|} = \alpha$ and t is a prefix of r with $|t| < |r|$ (t may be the empty word). For example $(1234)^3 = 123412341234$ and $(1234)^{\frac{7}{4}} = 1234123$. We say that a finite or infinite word w is α-*power-free* if w has no finite factors that are β-powers for $\beta \geq \alpha$. We say that a finite or infinite word w is α^+-power-free if w has no factors that are β-powers for $\beta > \alpha$, where $\alpha, \beta \geq 1$ are rational numbers. In some articles, it is a used convention to define that if a word w is "α-*power-free*" then α denotes a number or a "number with +"; see for instance [39]. We also apply this convention in our dissertation. The power-free words, also called repetition-free words, include well known square free (2-power-free), overlap free (2^+-power-free), and cube free words (3-power-free).

3 Enumeration

As mentioned in Introduction, in next subsections we discuss the enumeration of rich, closed, and privileged words and the factor complexity of rich words. First we introduce some more notation.

Let $\mathrm{Pal}_q \subseteq \mathrm{A}_q^*$ denote the set of all palindromes, let $\mathrm{Rich}_q \subseteq \mathrm{A}_q^*$ denote the set of all finite rich words, and let $\mathrm{Rich}_q^\infty \subseteq \mathrm{A}_q^\infty$ denote the set of all infinite rich words.

3.1 Rich words

Let $\Pi_q(n) = |\,\mathrm{Rich}_q \cap \mathrm{A}_q^n\,|$ be the number of rich words of length n. The enumeration of rich words is investigated in [44], where Vesti gives a recursive lower bound on the number of rich words of length n, and an upper bound on the number of binary rich words. Better lower and upper bounds can be found in [21], where Guo, Shallit, and Shur construct for each n a large set of rich words of length n. Consequently they prove that

$$\Pi_2(n) \geq \frac{C^{\sqrt{n}}}{p(n)}, \tag{1}$$

where $p(n)$ is a polynomial and C is a constant with $C \approx 37$. As already mentioned in Introduction, the language of rich words is factorial. It means that any factor of a rich word is also rich, see [20]. In particular it follows that

$$\Pi_q(n)\,\Pi_q(m) \geq \Pi_q(n+m)$$

for all positive integers m, n, q. Therefore, Fekete's lemma implies existence of the limit of $\sqrt[n]{\Pi_q(n)}$ and moreover

$$\lim_{n\to\infty} \sqrt[n]{\Pi_q(n)} = \inf\left\{ \sqrt[n]{\Pi_q(n)} \colon n \text{ is a positive integer} \right\}.$$

For a fixed n_0, one can find the number of all rich words of length n_0 and obtain an upper bound on the limit. Using a computer Rubinchik and Shur counted $\Pi_2(n)$ for $n \leq 60$ [35]. Since $\sqrt[60]{\Pi_2(60)} < 1.605$, in [21] the following upper bound is shown: $\Pi_2(n) \leq c1.605^n$ for some constant c.

In the article [[Ru02]] we show that $\Pi_q(n)$ has a subexponential growth on every finite alphabet. More precisely, we prove that

$$\lim_{n\to\infty} \sqrt[n]{\Pi_q(n)} = 1\,. \tag{2}$$

Our result is an argument in favor of a conjecture formulated in [21] saying that for some infinitely growing function $g(n)$ we have that

$$\Pi_2(n) = \mathcal{O}\Big(\frac{n}{g(n)}\Big)^{\sqrt{n}}. \tag{3}$$

To prove our result we use the following property of rich words [15, Definition 4 and Proposition 3]: the longest palindromic suffix of a rich word w has exactly one occurrence in w. We say that the longest palindromic suffix of w is *unioccurrent* in w. Using this property we prove that a rich word w can be factorized into p distinct nonempty palindromes. The crucial observation is that there is a constant c such that $p \leq c\frac{n}{\ln n}$, where $n = |w|$. We have that

$$\lim_{n\to\infty} \frac{c\frac{n}{\ln n}}{n} = 0.$$

This can be interpreted as follows: a rich word is a concatenation of a "small" number of distinct palindromes. Since every palindrome is determined by its first half, and since the language of rich words is a factorial language, we derive the subexponential upper bound on the number of rich words.

3.2 Factor Complexity of Rich words

Given a finite or infinite word $w \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$, we define the *factor complexity* of w as follows: $\mathrm{C}_w(n) = |\,\mathrm{F}_w(n)|$, where n is nonnegative integer. As mentioned in Introduction, the factor complexity $\mathrm{C}_w(n)$ enumerates the number of factors of length n in the word w.

It is well known that an infinite word $w \in \mathrm{A}_q^\infty$ is eventually periodic if and only if there exists a positive integer k such that $\mathrm{C}_w(k) \leq k$ [16]. It follows that for any aperiodic infinite word (not eventually periodic) it holds that $\mathrm{C}_w(n) \geq n+1$. The infinite words with factor complexity $\mathrm{C}_w(n) = n + 1$ are called Sturmian, [5, 6]. Sturmian words belong to extensively studied objects in combinatorics on words.

An infinite word $w \in \mathrm{A}_q^\infty$ is called *recurrent* if every factor $u \in \mathrm{F}_w$ has infinitely many occurrences in w. The word w is called *uniformly recurrent* if w is recurrent and for each factor $u \in \mathrm{F}_w$ there is an integer β_u such the distance between every two consecutive occurrences of u is bounded by β_u.

The palindromic complexity $\mathrm{D}_w(n)$ of the word $w \in \mathrm{A}_q^* \cup \mathrm{A}_q^\infty$ is defined as

$$\mathrm{D}_w(n) = |\,\mathrm{F}_w(n) \cap \mathrm{Pal}_q\,|.$$

In [1], it was shown for arbitrary infinite aperiodic word w that

$$\mathrm{D}_w(n) < \frac{16}{n}\,\mathrm{C}_w(n + \lfloor \frac{n}{4} \rfloor).$$

In [5], it is shown that if n is a positive integer and $w \in \mathrm{A}_q^\infty$ is a uniformly recurrent infinite word with $\mathrm{F}_w(n)$ closed under reversal, then

$$\mathrm{D}_w(n) + \mathrm{D}_w(n+1) \leq \mathrm{C}_w(n+1) - \mathrm{C}_w(n) + 2. \tag{4}$$

In [5] the authors proved the inequality (4) for uniformly recurrent words, but in the proof only "recurrent" is applied. Moreover, it is known that if F_w is closed under reversal, then w is recurrent [10, Proposition 2.2]. Thus the inequality (4) holds for every infinite word $w \in \mathrm{A}_q^\infty$ with F_w closed under reversal.

In [10] it was shown for rich words that the inequality (4) becomes equality; formally for every rich word w and a positive integer n we have that

$$\mathrm{D}_w(n) + \mathrm{D}_w(n+1) = \mathrm{C}_w(n+1) - \mathrm{C}_w(n) + 2. \tag{5}$$

The main result of our article [[Ru03]] states a quasi-polynomial upper bound for the factor complexity of rich words; more specifically we show that there are real constants c_1, c_2 such that for every rich word $w \in \mathrm{Rich}_q \cup \mathrm{Rich}_q^\infty$ and every positive integer n we have that

$$\mathrm{C}_w(n) \leq c_1 n^{c_2 \ln n}. \tag{6}$$

In addition we construct also an upper bound for the palindromic complexity of rich words and we prove the inequality (4) for finite words v whose set of factors $\mathrm{F}_v(n+1)$ is closed under reversal. Consequently we prove also the equality (5) for finite rich words w whose set of factors $\mathrm{F}_w(n+1)$ is closed under reversal. Then we apply the equality (5) to improve the upper bound for the factor and palindromic complexity of rich words.

To prove our results we use two properties of rich words. The second one uses the notion of a complete return. Given a word w and a factor r of w, we call the factor r a *complete return* to u in w if r contains exactly two occurrences of u, one as a prefix and one as a suffix; it follows that the complete return is a closed word. We state both properties as lemmas:

Lemma 1. *(see [11]) A factor r of a rich word w is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.*

Some generalizations of Lemma 1 may be found in [28].

Lemma 2. *(see [20]) Let w be a rich word. All complete returns to any palindromic factor u in w are palindromes.*

In [[Ru03]], we define a *switch* to be a factor of the form aub, where a, b are distinct letters and u is a palindrome. Applying Lemma 1, we show that the switch aub is uniquely determined by a, b and by the longest proper palindromic prefix of u. In addition, we prove that if the longest proper palindromic suffix of u is "too" long, then u is periodic with a "short" period. In consequence, we have that any switch is uniquely determined by two letters and a "short" palindrome. This allows us to present an upper bound for the number of switches of a given length.

Based on Lemma 2, we show that if a rich word w contains two palindromic factors xux and yuy, where x, y are distinct letters, then w has to contain a switch aub, where the letters a, b are not necessarily equal to the letters x, y. This observation allows to derive an upper bound for the palindromic complexity from the upper bound for the number of switches. From the palindromic complexity and Lemma 1, we prove our result for the factor complexity.

3.3 Closed and Priviledged Words

Let $\mathrm{Clo}_q \subseteq \mathrm{A}_q^*$ denote the set of all closed words and let $\mathrm{Priv}_q \subseteq \mathrm{A}_q^*$ denote the set of all privileged words.

Privileged words have been introduced quite recently in [23]. The combinatorial properties privileged words have been studied in [30, 38]. One of the questions that has been investigated was the enumeration of privileged words. In [27], it was proved that there are constants c and n_0 such that for all $n > n_0$, we have that

$$|\,\mathrm{Priv}_q \cap \mathrm{A}_q^n\,| \geq \frac{cq^n}{n(\log_q n)^2}. \tag{7}$$

The result (7) improves the lower bound for the number of privileged words from [17]. Since every privileged word is a closed word, the inequality (7) forms also a lower bound for the number of closed words.

Concerning an upper bound for the number of privileged words we have found the following open problem [29]: Give a nontrivial upper bound for the number of privileged words of length n. We have found no answers to this open problem.

In the article [[Ru06]] we show that

$$|\,\mathrm{Clo}_q \cap \mathrm{A}_q^n\,| \leq c \ln n \frac{q^n}{\sqrt{n}}, \tag{8}$$

where $n > 1$ and c is a some positive constant. Since privileged words are a subset of closed words, the formula (8) gives also an upper bound for the number of privileged words.

To prove our upper bound for the number of closed words, we split the language of closed words into words with a "short" and "long" border. For a closed word w with $|w| = n$, we define that w has a long border if w has a border u with $|u| \geq c \ln |w|$ for some predefined constant c. We derive an upper bound for the number of closed words with "long" border and we show the relation between the number of closed words with "short" and "long" borders. In consequence, we derive the upper bound (8).

4 Transition Property and Extendability

In 1985, Restivo and Salemi presented a list of five problems that deal with the question of transition property and extendability of power-free words [34]:

- Problem 1: Given an α-power-free word u, decide whether for every positive integer n there are words w, v such that $|w| = |v| = n$ and such that:
  - uv is α-power-free,
  - wu is α-power-free, and
  - wuv is α-power-free.
- Problem 2: Given an α-power-free word u, construct, if it exists, an infinite α-power-free word having u as a prefix.
- Problem 3: Given an arbitrary positive integer k, does there exists an α-power-free word u such that:
  - there exists a word v of length k such that uv is α-power-free and
  - for every word $\bar{v}$ with $|\bar{v}| > |v|$ we have that $u\bar{v}$ is not α-power-free.
- Problem 4: Given α-power-free words u and v, decide whether there is a transition word w, such that uwu is α-power-free.
- Problem 5: Given α-power-free words u and v, find a transition word w, if it exists.

Problems 1, 2, and 3 concern extendability of power-free words. Problems 4 and 5 concern transition property of power-free words. In general, these problems remain open. A recent survey on the progress of solving all the five problems can be found in [31]; in that article Petrova and Shur construct the transition words for cube-free words.

Although the problems are stated for power-free words, Problems 4 and 5 are also challenging for rich words. Let us compare rich and power-free words from the point of view of extendability. If w is a rich word, then there are letters x, y such that wx, yw, ywx are rich [44]. Thus every rich word can be extended. This property does not hold, in general, for power-free words. Thus Problems 1,2, and 3 are easy for rich words. However we elaborate for rich words some related questions concerning extendability [[Ru05]].

In next subsections we present our results related to transition property and extendability of power-free and rich words. In addition, we compare the properties of power-free and rich words from the point of view of transition property.

4.1 Transition property

Problems 4 and 5 of Restivo and Salemi are addressed in our article [[Ru07]], where we show for a wide variety of configurations (α, q) that for any right extendable α-power-free word u and any left extendable α-power-free word v over an alphabet A_q there is a transition word w such that uwv is also α-power-free over the alphabet A_q. We also construct the transition word w.

The very basic idea of our proof is that if u, v are α-power-free words and x is a letter such that x is not a factor of both u and v, then clearly uxv is α-power-free, provided that $\alpha \geq 2$. Note that there cannot be a factor in uxv which is an α-power and contains x, because x has only one occurrence in uxv. Less formally said, if u, v are α-power-free words over an alphabet with k letters, then we construct a "transition" word w over an alphabet with $k-1$ letters such that uwv is α-power-free. The proof involves intricate observations about recurrent factors of w.

In [28], the following open problem was stated.

- We do not know how to decide whether two rich words u and v are factors of a same rich word w.

This problem deals with transition property of rich words. In the article [[Ru04]] we show that if such w exists, then there is also a transition word $\overline{w}$ with a bounded length depending on the lengths of u and v. More exactly we show that: There are constants c_1, c_2 such that if w_1, w_2, w are rich words, $m = \max\{|w_1|, |w_2|\}$, and $\{w_1, w_2\} \subseteq \mathrm{F}_w$, then there exists a rich word $\overline{w}$ with $\{w_1, w_2\} \subseteq \mathrm{F}_{\overline{w}}$ and $|\overline{w}| \leq m2^{k(m)+2}$, where $k(m) = c_1 m^{c_2 \ln m}$. The constants c_1, c_2 depend on the size of the alphabet. Hence it is enough to check all rich words of length equal to or less than $m2^{k(m)+2}$ in order to decide if there is a rich word containing factors w_1, w_2.

Thus using a brute force we can decide the question of existence of a transition word and also we can construct a transition word. However this brute force algorithm is very inefficient and some significant improvement remains as an open question.

Comparing the quasi-polynomial upper bound for the factor complexity of rich words (6) with the lower bound for the number of binary rich words (1) we see that an infinite rich word contains only a "small" share of all rich words. Because of Lemma 1, this is not really surprising. Lemma 1 allows us to deduce that there are "many" pairs (u, v) of finite rich words that cannot be joined into a common rich word. For example consider the set $P = \{010111^i0110 \mid i \geq 1\}$. The set P contains words 0101110110, 01011110110, 010111110110, ... It is easy to see that if $u \in P$, then u is rich, the longest palindromic prefix of u is 010, and the longest palindromic suffix of u is 0110. Lemma 1 implies that if w is a rich word (finite or infinite) then $|\mathrm{F}_w \cap P| \leq 1$. In other words if $u, v \in P$ are distinct words, then there is no transition word w with uwv being rich.

It is interesting to note the contrast to power-free words, where an infinite α-power word can contain "almost" all extendable finite α-power words [39].

4.2 Extendability of rich words

We say that a rich word w *can be extended in at least two ways* if there are two distinct letters x, y such that wx, wy are rich. The extendability of rich words has been investigated in [44]. The author shows that if w is a rich word, then there is a word u with $|u| \leq 2|w|$ such that wu can be extended in at least two ways. It was presented as an open question to improve the upper bound for the length of u [44]. We address this open question in the article [[Ru05]].

We prove that if w is a finite rich word, then there is a rich word u with $|u| \leq |w|$ such that wu is a rich word that can be extended in at least two ways. In addition we investigate also a lower bound on the length of v. Given $w \in \mathrm{Rich}_q$, let $\mathrm{Rext}_q(w) \subseteq \mathrm{Rich}_q$ denote the set of all finite rich words such that if $u \in \mathrm{Rext}_q(w)$ then $wu \in \mathrm{Rich}_q$ and wu can be extended in at least two ways. Let

$$\omega_q(w) = \min\{|u| \mid u \in \mathrm{Rext}_q(w)\}$$

and let

$$\phi_q(n) = \max\{\omega_q(w) \mid w \in \mathrm{Rich}_q \text{ and } |w| = n\},$$

where $n > 0$. We prove that for each real constant $c > 0$ and each integer $m > 0$ there is $n > m$ such that

$$\phi_q(n) \geq \left(\frac{2}{9} - c\right) n. \tag{9}$$

The inequality (9) says that for each positive integer m and for each positive real c, there are an integer $n > m$ and a rich word $w \in \mathrm{Rich}_q$ with $|w| = n$ such that if u is a nonempty rich word and wu can be extended in at least two ways, then u is longer than $(\frac{2}{9} - c)|w|$. This can be formulated also as follows:

$$\limsup_{n \to \infty} \frac{\phi_q(n)}{n} \geq \frac{2}{9}.$$

5 Palindromic length

The palindromic length $\mathrm{PalLen}(v)$ of a finite word v is the minimal number of palindromes whose concatenation is equal to v. In 2013, Frid, Puzynina, and Zamboni presented the following conjecture.

Conjecture 1. *(see [18]) If w is an infinite word and k is an integer such that $\mathrm{PalLen}(u) \leq k$ for every factor u of w, then w is eventually periodic.*

In [18] the conjecture was proved for infinite words that are k-power-free for some positive integer k. It follows that if w is an infinite word with bounded palindromic length, then for each positive integer j there is a nonempty factor r such that r^j is a factor of w. Conjecture 1 attracted a lot of attention and there are quite a lot articles solving the conjecture for some classes of infinite words or investigating properties of palindromic length; see for instance [2, 3, 19, 35, 36]. However the conjecture still remains open.

In the article [[Ru08]] we bring some more insight in the infinite words with bounded palindromic length. Let w be an infinite word with bounded palindromic length. We show that for each positive integer j there are palindromes a, b with b nonempty such that $(ab)^j$ is a factor of w. Realize that $(ab)^i a$ is a palindromic factor of w for every $i < j$. In this way we can say that w contains many periodic palindromes. These results justify the following question: What is the palindromic length of a concatenation of a suffix of b and a periodic word $(ab)^j$ with "many" periodic palindromes? Our main result addresses this question.

In [36, Lemma 6] it was shown that if u, v are nonempty words then

$$|\operatorname{PalLen}(uv) - \operatorname{PalLen}(u)| \leq \operatorname{PalLen}(v).$$

In the article [[Ru08]] we show that if a, b are palindromes, b is nonempty, u is a nonempty suffix of b, $|ab|$ is the minimal period of aba, and j is a positive integer with $j \geq 3\operatorname{PalLen}(u)$, then

$$\operatorname{PalLen}(u(ab)^j) - \operatorname{PalLen}(u) \geq 0.$$

The proof is based on careful observations of some symmetries in the words with many periodic palindromes.

6 De Bruijn Sequences

In 1894, A. de Rivière formulated a question about existence of circular arrangements of 2^n zeros and ones in such a way that every word of length n appears exactly once, [14]. Let $B(n)$ denote the set of all such arrangements and let $B_0(n) \subseteq B(n)$ denote the elements that start with n zeros. It is easy to see that $|B_n| = 2^n |B_0(n)|$.

The question was solved in the same year by C. Flye Sainte-Marie, [37], together with presenting a formula for counting these arrangements:

$$|B_0(n)| = 2^{2^{n-1}-n}.$$

However the article was then forgotten. The topic became well known through the article of N.G. de Bruijn, who proved the same formula for the size of $B_0(n)$, [8]. Some time after, the article of C. Flye Sainte-Marie was rediscovered by Stanley, and it turned out that both proofs were principally the same, [9].

Stanley formulated in 2009 the following open problem [41], [42]:

> A binary de Bruijn sequence of degree n is a binary sequence $a_1 a_2 \cdots a_{2^n}$ (a_i is 0 or 1) such that all circular factors $a_i a_{i+1} \cdots a_{i+2^n-1}$ (taking subscripts modulo 2^n) of length n are distinct. An example of such sequence for $n = 3$ is 00010111. The number of binary de Bruijn

sequences of degree n is $2^{2^{n-1}}$. Let $B(n)$ denote the set of all binary de Bruijn sequences, then we want a bijection

$$\phi : B(n) \times B(n) \to \{0,1\}^{2^n}.$$

We solved this open problem of Stanley in the article [[Ru01]] by introducing new suitable bijections between de Bruijn graphs. Recall that a de Bruijn graph H_n is a directed graph with q^n nodes, whose nodes correspond to the words of length $n-1$ over an alphabet $\mathrm{A}_q = \{0, 1, \ldots, q-1\}$. A node $s_1 s_2 \ldots s_{n-1}$ has q outgoing edges to the nodes

$$s_2 \ldots s_{n-1}0,\ s_2 \ldots s_{n-1}1,\ \ldots,\ s_2 \ldots s_{n-1}(q-1).$$

It follows that a node $s_1 s_2 \ldots s_{n-1}$ has q incoming edges from nodes

$$0s_1 s_2 \ldots s_{n-2},\ 1s_1 s_2 \ldots s_{n-2},\ \ldots,\ (q-1)s_1 s_2 \ldots s_{n-2}.$$

De Bruijn graphs found several interesting applications, among others in networking, [4], and bioinformatics, [13, 32]. A Rauzy graph is a subgraph of the de Bruijn graph. In combinatorics of words, Rauzy graphs have been used for computing the factor complexity of infinite words [5, 6].

7 Dissecting of Infinite Languages

This section deals with some terms from language and automata theory that have not been formally defined in this dissertation; for example regular languages, context-free languages, deterministic finite automaton, and context free grammar. In case the reader is not familiar with automata and language theory, we refer to the book [22] or any other introduction book in that field.

In this section we consider also the alphabet with one letter; it means that q can be any positive integer.

An infinite language $L \subseteq \mathrm{A}_q^*$ is called *constantly growing* if there is a positive constant c_0 and a finite set K of positive integers such that for each $w \in L$ with $|w| \geq c_0$ there is a word $\bar{w} \in L$ and a constant $c \in K$ such that $|\bar{w}| = |w| + c$. We say also that L is (c_0, K)-constantly growing.

Given two infinite languages $L_1, L_2 \subseteq \mathrm{A}_q^*$, we say that L_1 *dissects* L_2 if

$$|L_1 \cap L_2| = \infty \quad \text{and} \quad |(\mathrm{A}_q^* \setminus L_1) \cap L_2| = \infty.$$

In [45], it has been proved that if L is a (c_0, K)-constantly growing language then there is a regular language $M \subseteq \mathrm{A}_q^*$ such that M dissects L and M is accepted by a deterministic finite automaton with $|K| + 1$ states.

In the article [[Ru09]] we define a *tetration* function (a repeated exponentiation) as follows:

$$\exp^{1,\alpha} = 2^\alpha$$

and

$$\exp^{j+1,\alpha} = 2^{\exp^{j,\alpha}},$$

where j, α are positive integers. The tetration function is known as a fast growing function of the argument α. If k, α are positive positive integers and $L \subseteq \mathrm{A}_q^*$ is an infinite language such that for each $u \in L$ there is $v \in L$ with $|u| < |v| \leq \exp^{k,\alpha}|u|$, then we call L a language with the *growth bounded by* (k, α)-tetration.

Given a context free language L, let $\kappa(L)$ denote the size of the smallest context free grammar G that generates L. We define the size of a grammar to be the total number of symbols on the right sides of all production rules.

In the article [[Ru09]] we show that if q, k are positive integers with $k \geq 2$, then there are context free languages $L_1, L_2, \dots, L_{3k-3} \subseteq \mathrm{A}_q^*$ with $\kappa(L_i) \leq 40k$ such that: If α is a positive integer and $L \subseteq \mathrm{A}_q^*$ is an infinite language with the growth bounded by (k, α)-tetration then there is a regular language M such that

$$M \cap \left(\bigcap_{i=1}^{3k-3} L_i \right)$$

dissects L and the minimal deterministic finite automaton accepting M has at most $k + \alpha + 3$ states.

We explain the basic idea of the proof. Recall that a non-associative word on the letter z is a "well parenthesized" word containing a given number of occurrences of z. For example for $n = 3$ occurences of z, the non-associative words are $(((zz)z)z)$, $((zz)(zz))$, $(z(z(zz)))$, $(z((zz)z))$, and $((z(zz))z)$. Every non-associative word contains the prefix $(^k z$ for some nonnegative integer k, where $(^k$ denotes the k-th power of the opening bracket. There are non-associative words such that k equals "approximately" $\log_2 n$. We construct three context free languages whose intersection accepts such words; we call these words *balanced non-associative words*. By counting the number of opening brackets of a balanced non-associative word with n occurrences of z we can compute a logarithm of n. By "chaining" this construction we are able to compute a repeated logarithm of n. This will allow us to transform the problem of dissecting of a language with the growth bounded by (k, α)-tetration to the problem of dissecting of a constantly growing language.

8 Future directions

Of course, in principle, it is possible to improve every result of our articles. However we suggest the following directions for future research. The selection of the directions is based mainly on experimental data that encourage us to believe that the future research in proposed themes would be rewarding.

8.1 Rich words

Although we proved the subexponential upper bound (2) for the number of rich words, the conjectured upper bound (3) remains open. One direction of the future research is to improve the upper bound in such a way to be able to confirm the conjectured upper bound (3) suggested by Guo, Shallit, and Shur. One possible approach would be to use the upper bound (6) for the factor complexity. It seems to us possible to apply the "low" factor complexity to derive some better upper bound for the number of rich words.

Concerning the quasi-polynomial upper bound for the factor complexity (6), it would be interesting to know if there is a polynomial upper bound; it means if there are positive constants c_1, c_2 such that $\mathrm{C}_w(n) \leq c_1 n^{c_2}$, where $n > 0$ and w is a rich word.

As already mentioned in Section 4.1, the practical application of our result from [[Ru04]] for finding a rich word containing two given factors requires a brute force, which turns out to be very inefficient. To improve our result or to improve the brute force algorithm remains as an open problem for the future research.

8.2 Power-free words

In 2009, Shur presented the following conjecture related to Problems 4 and 5 of Restivo and Salemi [40]:

Conjecture 2. *Let L be a power-free language and let $e(L) \subseteq L$ be the set of words of L that can be extended to a bi-infinite word respecting the given power-freeness. If $u, v \in e(L)$ then $uwv \in e(L)$ for some word w.*

In 2018, Conjecture 2 was presented also in [39] in a slightly different form.

We believe that our proof from [[Ru07]] could be generalized in order to confirm Conjecture 2 for α-power-free words with $\alpha > 2$ over an alphabet with $q \geq 3$ letters.

8.3 Palindromic length

For an infinite word w with bounded palindromic length, we identified factors u, v such that $\mathrm{PalLen}(uv) - \mathrm{PalLen}(v) \geq 0$ [[Ru08]]. The idea for the future development of this result is, for given positive integer k, to identify factors u, v of w such that $\mathrm{PalLen}(u) = k$ and $\mathrm{PalLen}(uv) - \mathrm{PalLen}(u) > 0$. The existence of such factors would, in consequence, allow us to prove Conjecture 1.

References

[1] J.-P. Allouche, M. Baake, J. Cassaigne, and D. Damanik, *Palindrome complexity*, Theoretical Computer Science, 292 (2003), pp. 9–31.

[2] P. Ambrož, O. Kadlec, Z. Masáková, and E. Pelantová, *Palindromic length of words and morphisms in class P*, Theoretical Computer Science, 780 (2019), pp. 74–83.

[3] P. Ambrož and E. Pelantová, *On palindromic length of Sturmian sequences*, in Developments in Language Theory, P. Hofman and M. Skrzypczak, eds., Cham, 2019, Springer International Publishing, pp. 244–250.

[4] J. Baker, *De Bruijn graphs and their applications to fault tolerant networks*, Oregon State University (2011).

[5] P. Baláži, Z. Masáková, and E. Pelantová, *Factor versus palindromic complexity of uniformly recurrent infinite words*, Theoretical Computer Science, 380 (2007), pp. 266–275.

[6] L. Balková, E. Pelantová, and Š. Starosta, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO-Theor. Inf. Appl., 44 (2010), pp. 443–470.

[7] J. Berstel and D. Perrin, *The origins of combinatorics on words*, European Journal of Combinatorics, 28 (2007), pp. 996–1022.

[8] N. G. D. Bruijn, *A combinatorial problem*, Nederl. Akad. Wetensch., 49 (1946), pp. 758–764.

[9] ——, *Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2n zeros and ones that show each n-letter word exactly once*, TH-Report 75-WSK-06, Technological University Eindhoven, (1975), pp. 1–14.

[10] M. Bucci, A. De Luca, A. Glen, and L. Q. Zamboni, *A connection between palindromic and factor complexity using return words*, Adv. in Appl. Math., 42 (2009), pp. 60–74.

[11] ——, *A new characteristic property of rich words*, Theoretical Computer Science, 410 (2009), pp. 2860–2863.

[12] J. D. Currie and L. Mol, *The undirected repetition threshold*, in Combinatorics on Words, R. Mercaş and D. Reidenbach, eds., Cham, 2019, Springer International Publishing, pp. 145–158.

[13] E. B. Daniel R. Zerbino, *Velvet: Algorithms for de novo short read assembly using de Bruijn graphs*, Genome Research 18 (5), (2008), pp. 821–829.

[14] A. de Rivière, *Question nr. 48*, l'Intermédiaire des Mathématiciens, (1894), pp. 19–20.

[15] X. Droubay, J. Justin, and G. Pirillo, *Episturmian words and some constructions of de Luca and Rauzy*, Theoretical Computer Science, 255 (2001), pp. 539–553.

[16] N. P. Fogg, *Substitutions in Arithmetics, Dynamics and Combinatorics*, vol. 1794 of Lecture notes in mathematics, Springer, 1st ed., 2002.

[17] M. Forsyth, A. Jayakumar, J. Peltomäki, and J. Shallit, *Remarks on privileged words*, International Journal of Foundations of Computer Science, Vol. (27), No. 04, (2016), pp. 431–442.

[18] A. Frid, S. Puzynina, and L. Zamboni, *On palindromic factorization of words*, Advances in Applied Mathematics, 50 (2013), pp. 737–748.

[19] A. E. Frid, *Sturmian numeration systems and decompositions to palindromes*, European Journal of Combinatorics, 71 (2018), pp. 202–212.

[20] A. Glen, J. Justin, S. Widmer, and L. Q. Zamboni, *Palindromic richness*, European Journal of Combinatorics, 30 (2009), pp. 510–531.

[21] C. Guo, J. Shallit, and A. M. Shur, *Palindromic rich words and run-length encodings*, Information Processing Letters, 116 (2016), pp. 735–738.

[22] J. Hopcroft, R. Motwani, and J. Ullman, *Introduction to Automata Theory, Languages, an Computation*, Addison Wesley, 2001.

[23] J.Kellendonk, D.Lenz, and J.Savinien, *A characterization of subshifts with bounded powers*, Discrete Mathematics, Volume 313, Issue 24, (2013), pp. 2881–2894.

[24] M. Lothaire, *Combinatorics on Words*, vol. 17 of Encyclopaedia of Mathematics and its Applications, Addison-Wesley, Reading, Mass., 1983. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, Cambridge UK, 1997.

[25] ——, *Algebraic Combinatorics on Words*, no. 90 in Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2002.

[26] T. Ng, P. Ochem, N. Rampersad, and J. Shallit, *New results on pseudosquare avoidance*, in Combinatorics on Words, R. Mercaş and D. Reidenbach, eds., Cham, 2019, Springer International Publishing, pp. 264–274.

[27] J. Nicholson and N. Rampersad, *Improved estimates for the number of privileged words*, Journal of Integer Sequences, 21 (2018).

[28] E. Pelantová and Š. Starosta, *On words with the zero palindromic defect*, in Combinatorics on Words, S. Brlek, F. Dolce, C. Reutenauer, and É. Vandomme, eds., Cham, 2017, Springer International Publishing, pp. 59–71.

[29] J. Peltomäki, *Privileged words and Sturmian words*, Turku Centre for Computer Science, TUCS Dissertations No 214, August 2016.

[30] ——, *Introducing privileged words: Privileged complexity of Sturmian words*, Theoretical Computer Science, Volume 500, (2013), pp. 57–67.

[31] E. A. Petrova and A. M. Shur, *Transition property for cube-free words*, in Computer Science – Theory and Applications, R. van Bevern and G. Kucherov, eds., Cham, 2019, Springer International Publishing, pp. 311–324.

[32] P. A. Pevzner, H. Tang, and M. S. Waterman, *An Eulerian path approach to DNA fragment assembly*, Proc. Natl. Acad. Sci., 98 (2001), pp. 9748–9753.

[33] S. Puzynina, *Abelian properties of words*, in Combinatorics on Words, R. Mercaş and D. Reidenbach, eds., Cham, 2019, Springer International Publishing, pp. 28–45.

[34] A. Restivo and S. Salemi, *Some decision results on nonrepetitive words*, in Combinatorial Algorithms on Words, A. Apostolico and Z. Galil, eds., Berlin, Heidelberg, 1985, Springer Berlin Heidelberg, pp. 289–295.

[35] M. Rubinchik and A. M. Shur, *EERTREE: An Efficient Data Structure for Processing Palindromes in Strings*, Springer International Publishing, Cham, 2016, pp. 321–333.

[36] A. Saarela, *Palindromic length in free monoids and free groups*, in Combinatorics on Words, S. Brlek, F. Dolce, C. Reutenauer, and É. Vandomme, eds., Cham, 2017, Springer International Publishing, pp. 203–213.

[37] C. F. Sainte-Marie, *Solution to question nr. 48*, l'Intermédiaire des Mathématiciens, (1894), pp. 107–110.

[38] L. Schaeffer and J. Shallit, *Closed, palindromic, rich, privileged, trapezoidal, and balanced words in automatic sequences*, Electronic Journal of Combinatorics, 23 (2016), p. P1.25.

[39] J. Shallit and A. Shur, *Subword complexity and power avoidance*, Theoretical Computer Science, 792 (2019), pp. 96–116. Special issue in honor of the 70th birthday of Prof. Wojciech Rytter.

[40] A. M. Shur, *Two-sided bounds for the growth rates of power-free languages*, in Developments in Language Theory, V. Diekert and D. Nowotka, eds., Berlin, Heidelberg, 2009, Springer Berlin Heidelberg, pp. 466–477.

[41] R. P. Stanley, *Bijective proof problems, problem 28*, MIT Mathematics Department, available at `http://www-math.mit.edu/~rstan/bij.pdf`, (2009).

[42] ——, *Topics in algebraic combinatorics*, MIT Mathematics Department, available at `http://www-math.mit.edu/~rstan/algcomb/algcomb.pdf`, (2009).

[43] A. Thue, *Über unendliche Zeichenreihen*, Christiana Videnskabs-Selskabs Skrifter,I. Math.-naturv. Klasse 7, (1906).

[44] J. Vesti, *Extensions of rich words*, Theoretical Computer Science, 548 (2014), pp. 14–24.

[45] T. Yamakami and Y. Kato, *The dissecting power of regular languages*, Information Processing Letters, 113 (2013), pp. 116–122.

Article [[Ru01]]: Bijections in de Bruijn Graphs

Bijections in de Bruijn Graphs

Josef Rukavicka*

December 10, 2016
Mathematics Subject Classification: 05C30, 05A19
Keywords: De Bruijn Graph, De Bruijn Sequence, T-net, Bijection

Abstract

A T-net of order m is a graph with m nodes and $2m$ directed edges, where every node has indegree and outdegree equal to 2. (A well known example of T-nets are de Bruijn graphs.) Given a T-net N of order m, there is the so called "doubling" process that creates a T-net N^* from N with $2m$ nodes and $4m$ edges. Let $|X|$ denote the number of Eulerian cycles in a graph X. It is known that $|N^*| = 2^{m-1}|N|$. In this paper we present a new proof of this identity. Moreover we prove that $|N| \leq 2^{m-1}$.
Let $\Theta(X)$ denote the set of all Eulerian cycles in a graph X and $S(n)$ the set of all binary sequences of length n. Exploiting the new proof we construct a bijection $\Theta(N) \times S(m-1) \to \Theta(N^*)$, which allows us to solve one of Stanley's open questions: we find a bijection between de Bruijn sequences of order n and $S(2^{n-1})$.

1 Introduction

In 1894, A. de Rivière formulated a question about existence of circular arrangements of 2^n zeros and ones in such a way that every word of length n appears exactly once, [7]. Let $B_0(n)$ denote the set of all such arrangements. (we apply the convention that the elements of $B_0(n)$ are binary sequences that start with n zeros). The question was solved in the same year by C. Flye Sainte-Marie, [5], together with presenting a formula for counting these arrangements: $|B_0(n)| = 2^{2^{n-1}-n}$. However the paper was then forgotten. The topic became well known through the paper of N.G. de Bruijn, who proved the same formula for the size of $B_0(n)$, [2]. Some time after, the paper of C. Flye Sainte-Marie was rediscovered by Stanley, and it turned out that both proofs were principally the same, [3].

*Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, CZECH TECHNICAL UNIVERSITY IN PRAGUE (josef.rukavicka@seznam.cz).

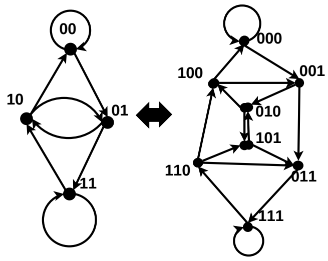

Figure 1: A doubling of a de Bruijn graph: N and N^*

The proof uses a relation between $B_0(n)$ and the set of Eulerian cycles in a certain type of T-nets: A T-net N of order m is defined as a graph with m nodes and $2m$ directed edges, where every node has indegree and outdegree equal to 2 (a T-net is often referred as a balanced digraph with indegree and outdegree of nodes equal to 2, see for example [10]). N.G. de Bruijn defined a doubled T-net N^* of N. A doubled T-net N^* of N is a T-net such that:

- each node of N^* corresponds to an edge of N
- two nodes in N^* are connected by an edge if their corresponding edges in N are incident and the ending node of one edge is the starting node of the second edge.

Remark We call two edges to be incident if they share at least one common node; the orientation of edges does not matter.

As a result N^* has $2m$ nodes and $4m$ edges, see an example on Figure 1. (A doubled T-net of N is known as well as a line graph of N, [4].)

Let $\Theta(X)$ be the set of all Eulerian cycles in X and let $|X| = |\Theta(X)|$ denote the number of Eulerian cycles in X, where X is a graph. It was proved inductively that $|N^*| = 2^{m-1}|N|$. Moreover N.G. de Bruijn constructed a T-net (nowadays called a de Bruijn graph) whose Eulerian cycles are in bijection with the elements of $B_0(n)$.

A de Bruijn graph H_n of order n is a T-net of order 2^n, whose nodes correspond to the binary words of length $n-1$. A node $s_1s_2\dots s_{n-1}$ has two outgoing edges to the nodes $s_2\dots s_{n-1}0$ and $s_2\dots s_{n-1}1$. It follows that a node $s_1s_2\dots s_{n-1}$ has two incoming edges from nodes $0s_1s_2\dots s_{n-2}$ and $1s_1s_2\dots s_{n-2}$. Given an edge e going from the node $s_1s_2\dots s_{n-1}$ to the node $s_2\dots s_{n-1}s_n$, then the edge e corresponds to the word $s_1s_2\dots s_{n-1}s_n$ of length n, which implies the natural bijection between Eulerian cycles $\Theta(H_n)$ and binary sequences $B_0(n)$, [2]. That is why we will write $B_0(n) \equiv \Theta(H_n)$.

De Bruijn graphs found several interesting applications, among others in networking, [1], and bioinformatics, [6], [8].

The important property of de Bruijn graphs is that a doubled T-net of a de Bruijn graph of order n is a de Bruijn graph of order $n+1$, see an example on Figure 1 of the de Bruijn graph of order 3 ($H_3 = N$) and of order 4 ($H_4 = N^*$). Since $|B_0(2)| = 1$ ($B_0(2) = \{0011\}$) it has been derived that $|B_0(n)| = 2^{2^{n-1}-n}$, [1], [2], [3].

There is also another proof using matrix representation of graphs, [10]. Yet it was an open question of Stanley, [9], [10], if there was a bijective proof:

> Let $B(n)$ be the set of all binary de Bruijn sequences of order n, and let $S(n)$ be the set of all binary sequences of length n. Find an explicit bijection $B(n) \times B(n) \to S(2^n)$.

This open question was solved in 2009, [4], [10].

Remark In the open question of Stanley, $B(n)$ denotes the de Bruijn sequences that do not necessarily start with n zeros like in the case of B_0. $B(n)$ contains all 2^n "circular rotations" of all sequences from $B_0(n)$; formally, given $s = s_1 s_2 \dots s_{2^n} \in B_0(n)$, then $s_i s_{i+1} \dots s_{2^n} s_1 s_2 \dots s_{i-1} \in B(n)$, where $1 \le i \le 2^n$. It is easy to see that all these 2^n "circular rotations" are distinct binary sequences. It follows that $|B(n)| = 2^n |B_0(n)|$. Hence it is enough to find a bijection $B_0(n) \to S(2^{n-1} - n)$ to solve this open question.

In this paper we present a new proof of the identity $|N^*| = 2^{m-1}|N|$, which allows us to prove that $|N| \le 2^{m-1}$ and to construct a bijection $\nu : \Theta(N) \times S(m-1) \to \Theta(N^*)$ and consequently to present another solution to the Stanley's open question: We define $\rho_2(\varepsilon) = 0011$ (recall that $B_0(2) = \{0011\}$) and let $\rho_n : S(2^{n-1} - n) \to B_0(n)$ be a map defined as $\rho_n(s) = \nu(\rho_{n-1}(\dot{s}), \ddot{s})$, where ε is the binary sequence of length 0, $n > 2$, $s = \dot{s}\ddot{s}$, $\dot{s} \in S(2^{n-2} - (n-1))$, and $\ddot{s} \in S(2^{n-2} - 1)$.

Proposition 1.1 *The map ρ_n is a bijection.*

Proof Note that $\dot{s} \in S(2^{n-2} - (n-1))$ and $|B_0(n-1)| = 2^{(n-1)-1} - (n-1) = 2^{n-2} - (n-1)$; thus $\dot{s}$ is a valid input for the function ρ_{n-1} and $\rho_{n-1}(\dot{s}) \in B_0(n-1) \equiv \Theta(H_{n-1})$. In addition, H_{n-1} has $m = 2^{n-2}$ nodes and $\ddot{s} \in S(2^{n-2} - 1)$ has the length $m-1$, hence it makes sense to define $\rho_n(s) = \nu(\rho_{n-1}(\dot{s}), \ddot{s})$. Because ν is a bijection, see Proposition 3.1, it is easy to see by induction on n that ρ_n is a bijection as well.

Remark Less formally said, the bijection $\rho_n(s)$ splits the binary sequence s into two subsequences $\dot{s}$ and $\ddot{s}$. Then the bijection ρ_{n-1} is applied to $\dot{s}$, the result of which is a de Bruijn sequence p from $B_0(n-1)$ (and thus an Eulerian cycle in H_{n-1}). Then the bijection ν is applied to p and $\ddot{s}$. The result is a de Bruijn sequence from $B_0(n)$.

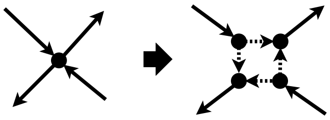

Figure 2: A node replacing by 4 nodes and 4 edges

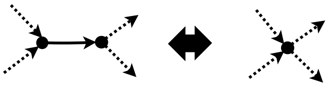

Figure 3: A removing solid edges and fusion of nodes

2 A double and quadruple of a T-net

Let Y be a set of graphs; we define $\Theta(Y) = \bigcup_{X \in Y} \Theta(X)$ (the union of sets of Eulerian cycles in graphs from Y) and $|Y| = \sum_{X \in Y} |X|$ (the sum of the numbers of Eulerian cycles). Let $U(X)$ denote the set of nodes of a graph X.

We present a new way of constructing a doubled T-net, which will enable us to show a new non-inductive proof of the identity $|N^*| = 2^{m-1}|N|$ and to prove $|N| \leq 2^{m-1}$.

We introduce a quadruple of N denoted by $\hat{N}$: The quadruple $\hat{N}$ arises from N by replacing every node $a \in U(N)$ by 4 nodes and 4 edges as depicted on the Figure 2. Let $\Gamma(a)$ denote the set of these 4 nodes and $\Pi(a)$ denote the set of these 4 edges that have replaced the node a. The edges from $\Pi(a)$ are dashed on the figures and we will distinguish dashed and solid edges as follows: In a graph containing at least one dashed edge, we define a *dashed Eulerian cycle* to be a cycle that traverses all dashed edges exactly once and all solid edges exactly twice, see Figure 4. In a graph without dashed edges, we define a *dashed Eulerian cycle* to be the same as an “ordinary” Eulerian cycle.

Remark Note that a quadruple $\hat{N}$ is not a T-net, since the indegree and outdegree are not always equal to 2. But since the solid edges can be traversed twice, we can consider them as parallel edges (two edges that are incident to the same two nodes). Then it would be possible to regard $\hat{N}$ as a T-net.

By removing solid edges and “fusing” their incident nodes into one node in $\hat{N}$ (as depicted on Figure 3), we obtain a doubled T-net N^* of N. And the reverse process yields $\hat{N}$ from N^*: turn all edges from solid to dashed and then replace every node by two nodes connected by one solid edge, where one node has two outgoing dashed edges and one incoming solid edge and the second node two incoming edges and one outgoing solid edge. Thus we have a natural bijection between *dashed Eulerian cycles* in $\hat{N}$ and N^*. See an example on Figure 4.

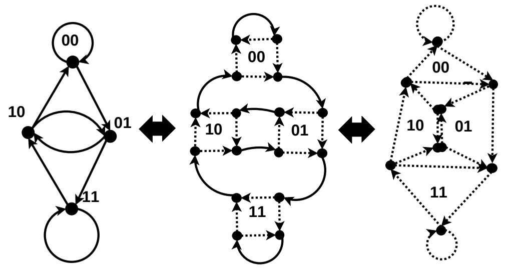

Figure 4: An example of N, $\hat{N}$, and N^*

Remark If all edges in a graph are solid or if all edges in a graph are dashed, then it makes no difference if they are solid or dashed. A *dashed Eulerian cycle* traverses in that case just once every edge.

Fix an order on nodes $U(N)$. As a result we have a bijection $\phi : \{1,2,\ldots,m\} \to U(N)$. Given $i \in \{1,2,\ldots,m\}$, let us denote the edges from $\Pi(\phi(i))$ by t,u,v,z, in such a way that t and v are not incident edges; it follows that u and z are not incident as well.

Let $W_0 = \{\hat{N}\}$, we define $W_i = \{\dot{w}, \ddot{w} \mid w \in W_{i-1}\}$, where $i \in \{1,2,\ldots,m\}$ and $\dot{w}$, $\ddot{w}$ are defined as follows: We construct the graph $\dot{w}$ by removing edges t,v from w and by changing the edges u,z from dashed to solid (thus allowing the edges u,z to be traversed twice). Similarly we construct $\ddot{w}$ from w by removing edges u,z and by changing t,v from dashed to solid, where $t,u,v,z \in \Pi(\phi(i))$.

The crucial observation is:

Proposition 2.1 *Let $w \in W_i$, where $i \in \{0,1,\ldots,m-2\}$. Then* $|w| = 2|\dot{w}| + 2|\ddot{w}|$.

Remark The following proof is basically identical to the one in [2], where the author constructed two graphs d_1, d_2 from a graph d and proved that $|d| = 2|d_1| + 2|d_2|$

Proof Given a *dashed Eulerian cycle* g in w, then split g in four paths A,B,C,D and edges $t,u,v,z \in \Pi(\phi(i))$. We will count the number of *dashed Eulerian cycles* in $\dot{w}, \ddot{w}$ that are composed from all 4 paths A,B,C,B and that differ only in their connections on edges t,u,v,z. Exploiting the N.G. de Bruijn's notation, all possible cases are depicted on Figures 5 and 6.

- In case I, the graph w contains 4 *dashed Eulerian cycles*: *AtBzDuCv*, *AtCuBzDv*, *AtCvDuBz*, *AzDuBtCv*; whereas the graphs $\dot{w}$ and $\ddot{w}$ have together 2 *dashed Eulerian cycles*: *AzDuCuBz*, *AtBtCvDv*. Thus $|w| = 4$ and $|\dot{w}|$+$|\ddot{w}| = 2$.

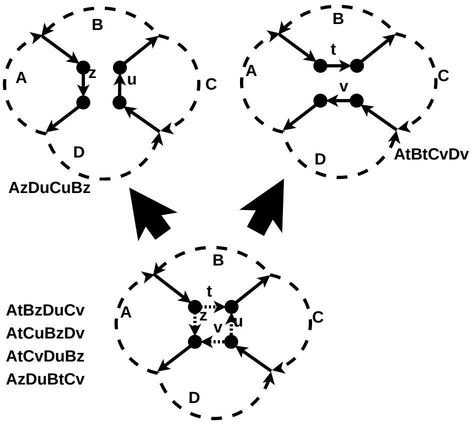

Figure 5: Edges replacement. Case I

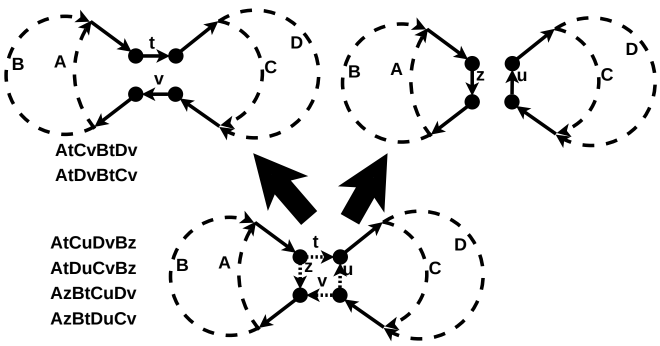

Figure 6: Edges replacement. Case II

- In case II, the graph w contains 4 *dashed Eulerian cycles*: $AtCuDvBz$, $AtDuCvBz$, $AzBtCuDv$, $AzBtDuCv$; whereas the graph $\ddot{w}$ has 2 *dashed Eulerian cycles*: $AtCvBtDv$, $AtDvBtCv$. The graph $\dot{w}$ is disconnected and therefore $\dot{w}$ has 0 *dashed Eulerian cycles*. Thus $|w| = 4$ and $|\dot{w}|+|\ddot{w}| = 2$. In case II, it is possible the $A = B$ or $C = D$. In such a case, $|w| = 2$ and $|\dot{w}|+|\ddot{w}| = 1$.

This ends the proof.

We define $\Delta = \{w \mid w \in W_m$ and w is connected$\}$. The Figure 7 shows an example of all iterations and construction of graphs in Δ from the graph $\hat{N}$, where N is a de Bruijn graph of order 3. The order of nodes from N is $00 < 10 < 01 < 11$. Most of the disconnected graphs are ommited.

Remark In the previous proof in case II, it can happen that $A = B$ or $C = D$. Note in the iteration step $i = m$ (when constructing W_m from W_{m-1}) it holds that $A = B$ and $C = D$, because all nodes have indegree and outdegree equal to 1 with exception of nodes $\Gamma(\phi(m-1))$. Hence $|W_{m-1}| = |W_m|$. It follows as well that every connected graph $w \in W_{m-1}$ has exactly one *dashed Eulerian cycle*. That is why in the Proposition 2.1 we consider $i \in \{0, 1, \ldots, m-2\}$.

Corollary 2.1 $2|W_{i-1}| = |W_i|$ *and* $|W_{m-1}| = |W_m|$*, where* $i \in \{1, 2, \ldots, m-1\}$.

Proposition 2.2 $2^{m-1}|\Delta| = |N^*| = |\hat{N}|$.

Proof The only graphs in W_m that contain a *dashed Eulerian cycle* are connected graphs, it means only graphs from Δ. On the other hand every graph $w \in \Delta$ contains exactly one *dashed Eulerian cycle*, since every node has indegree and outdegree equal to 1. The proposition follows then from Corollary 2.1, because $|\hat{N}| = |W_0|$ (recall that $W_0 = \{\hat{N}\}$).

Proposition 2.3 *There is a bijection between* $\Theta(N)$ *and* $\Theta(\Delta)$ *and* $\Theta(W_{m-1})$ *and* $\Theta(W_m)$.

Proof Given a connected graph $w \in W_{m-1}$, then just one graph of $\dot{w}$ and $\ddot{w}$ is connected. Let us say it is $\dot{w}$. Recall that there is exactly one *dashed Eulerian cycle* $AtCuCvAz$ in w ($A = B$ and $C = D$, see Figure 6). Then $AtCv$ is the only *dashed Eulerian cycle* in $\dot{w} \in \Delta \subset W_m$. This shows a bijection between $\Theta(W_{m-1})$ and $\Theta(W_m)$ and $\Theta(\Delta)$.

Let $\bar{p} = p_1 p_2 \ldots p_{4m}$ be the only *dashed Eulerian cycle* in $w \in \Delta$, where p_i are edges of w. Without loss of generality suppose that $p_1 \in \Pi(a)$ for some $a \in U(N)$ (it means that p_1 is a dashed edge in $\hat{N}$). It follows that all p_i with i odd are dashed edges in $\hat{N}$ all p_i with i even are edges from N (they are solid edges in $\hat{N}$); in consequence the path $p = p_2 p_4 \ldots p_{4m}$ is a *dashed Eulerian cycle* in N. A turning the *dashed Eulerian cycle* in w into the *dashed Eulerian cycle* p in N is schematically depicted on Figure 8. Thus we have a bijection between $\Theta(N)$ and $\Theta(\Delta)$. This ends the proof.

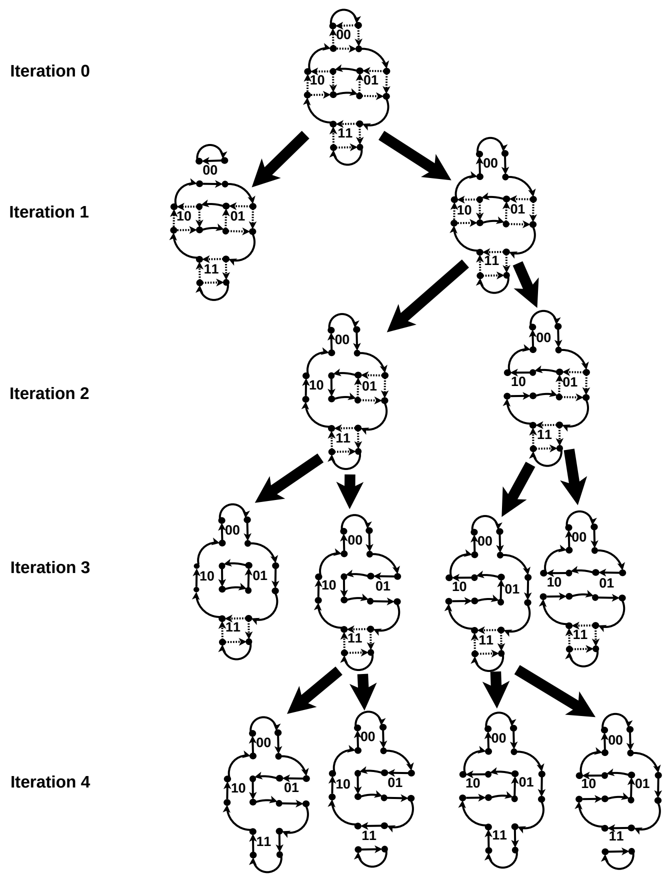

Figure 7: Constructing the set Δ from $\hat{N}$

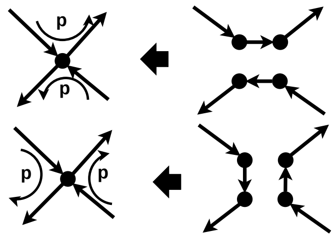

Figure 8: Converting a *dashed Eulerian cycle* from Δ into a *dashed Eulerian cycle* in N

Corollary 2.2 *Let N be a T-net of order m. Then* $|N| \leq 2^{m-1}$ dashed Eulerian cycles.

Proof The set W_{m-1} contains 2^{m-1} graphs and recall that every connected graph $w \in W_{m-1}$ has exactly one *dashed Eulerian cycle*. The result follows then from $|W_{m-1}| = |W_m|$ and $\Delta \subseteq W_m$.

3 Bijection of binary sequences and de Bruijn sequences

Given $i \in \{1, 2, \ldots, m\}$, in the previous section we agreed that the edges from $\Pi(\phi(i))$ are denoted by t, u, v, z, in such a way that t and v are not incident edges (and consequently that u and z are not incident as well). For this section we need that these edges are ordered, hence let us suppose that it holds $t < u < v < z$. This will allow us to identify "uniquely" the edges.

Let us look again on the Figure 5. We can identify the path A as the path between incident nodes of the edge z that do not contain edges t, u, v. In a similar way we can identify B, C, D.

On the Figure 6 we can not distinguish A from B and C from D only by edges t, u, v, z. If $A \neq B$, then let δ be the first node where A and B differ. The node δ has two outgoing dashed edges, let us say they are t, z. We use this difference to distinguish A and B. Let us define A to be the path that follows the edge t from δ and B the path that follows the edge z from δ. Again in a similarly way we can distinguish C from D. Hence let us suppose we have an "algorithm" that splits a *dashed Eulerian cycle* $p \in \Theta(W_i)$ into the paths A, B, C, D and edges $t, u, v, z \in \Pi(\phi(i))$ for given N, i (recall that the nodes of N are ordered and thus i determines the node $\phi(i) \in U(N)$). We introduce the function $\omega_{N,i} : (p, \alpha) \to \Theta(W_{i-1})$, where

- N is a T-net of order m

- $i \in \{1, \dots, m-1\}$
- $p \in \Theta(W_i)$
- $\alpha \in \{0, 1\}$

Remark Less formally said, the function ω transform a *dashed Eulerian cycle* $p \in \Theta(W_i)$ into a *dashed Eulerian cycle* $\bar{p} \in \Theta(W_{i-1})$ for given N, i, α.

Given N and i, we define for the case I (Figure 5):
$\omega_{N,i}(AzDuCuBz, 0) = AtBzDuCv$
$\omega_{N,i}(AzDuCuBz, 1) = AtCuBzDv$
$\omega_{N,i}(AtBtCvDv, 0) = AtCvDuBz$
$\omega_{N,i}(AtBtCvDv, 1) = AzDuBtCv$
For the case II (Figure 6), where $A \neq B$ and $C \neq D$:
$\omega_{N,i}(AtCvBtDv, 0) = AtCuDvBz$
$\omega_{N,i}(AtCvBtDv, 1) = AzBtCuDv$
$\omega_{N,i}(AtDvBtCv, 0) = AtDuCvBz$
$\omega_{N,i}(AtDvBtCv, 1) = AzBtDuCv$
For the case II where $A = B$ and $C \neq D$:
$\omega_{N,i}(AtCvAtDv, 0) = AtCuDvAz$
$\omega_{N,i}(AtCvAtDv, 1) = AtDuCvAz$
For the case II where $A \neq B$ and $C = D$:
$\omega_{N,i}(AtCvBtCv, 0) = AtCuCvBz$
$\omega_{N,i}(AtCvBtCv, 1) = AzBtCuCv$
Now, when we fixed an order on edges at the beginning of this section, it is necessary to distinguish another alternative of the case II, namely the paths A, B can be paths between incident nodes of the edge t that do not contain edges u, v, z and C, D can be paths between incident nodes of the edge v that do not contain edges t, u, z, let us denote it as case III, see Figure 9. We define ω in a similar way as for the case II:

For the case III (Figure 9), where $A \neq B$ and $C \neq D$:
$\omega_{N,i}(AuCzBuDz, 0) = AuCvDzBt$
$\omega_{N,i}(AuCzBuDz, 1) = AtBuCvDz$
$\omega_{N,i}(AuDzBuCz, 0) = AuDvCzBt$
$\omega_{N,i}(AuDzBuCz, 1) = AtBuDvCz$
For the case III where $A = B$ and $C \neq D$:
$\omega_{N,i}(AuCzAuDz, 0) = AuCvDzAt$
$\omega_{N,i}(AuCzAuDz, 1) = AuDvCzAt$
For the case III where $A \neq B$ and $C = D$:
$\omega_{N,i}(AuCzBuCz, 0) = AuCvCzBt$
$\omega_{N,i}(AuCzBuCz, 1) = AtBuCvCz$

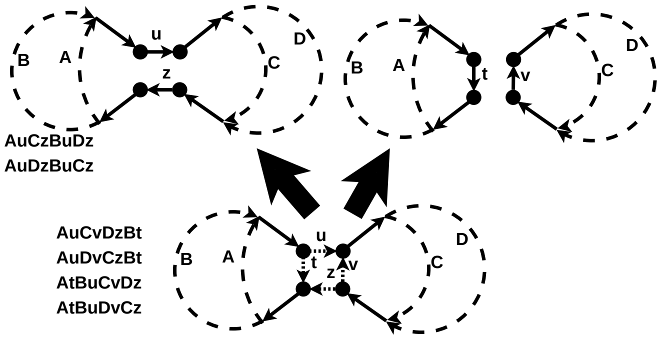

Figure 9: Edges replacement. Case III

Remark The previous definition of $\omega_{N,i}(p,\alpha)$ can be modified with regard to the reader's needs, including the way of recognition of paths A,B,C,D. It matters only that $\omega_{N,i}$ is injective. Our definition is just one possible way.

Remark To understand correctly the definition of ω, recall that when comparing two *dashed Eulerian cycles*, it does not matter which edge is written as the first one. For example the paths *AtCuDvAz* and *AzAtCuDv* are an identical *dashed Eulerian cycle*.

Let $S(n)$ denote the set of all binary sequences of length n.

Proposition 3.1 *Let N be a T-net of order m, $s = s_1 s_2 \dots s_{m-1} \in S(m-1)$ be a binary sequence, and $p \in \Theta(N)$. We define $p = p^{m-1}$ and $p^{i-1} = \omega_{N,i}(p^i, s_i)$, where $i \in \{1,2,\dots,m-1\}$. Then the map $\nu : \Theta(N) \times S(m-1) \to \Theta(N^*)$ defined as $\nu(p,s) = p^0$ is a bijection.*

Proof Recall that there is a bijection between $\Theta(N)$ and $\Theta(W_{m-1})$, see Proposition 2.3; hence we can suppose that $p \in W_{m-1}$.
The definition of the function ω implies that $\omega_{N,i}(p,\alpha) = \omega_{N,i}(\bar{p},\bar{\alpha})$ if and only if $p = \bar{p}$ and $\alpha = \bar{\alpha}$. It follows that ν is injective. In addition we proved that $|N| = |W_{m-1}|$ and that $2^{m-1}|N| = |\hat{N}| = |W_0|$. In consequence ν is surjective and thus bijective.

References

[1] J. Baker: *De Bruijn graphs and their applications to fault tolerant networks*, 2011, `http://hdl.handle.net/1957/26215`

[2] N. G. De Bruijn: *A combinatorial problem*, Nederl. Akad. Wetensch., Proc., 49:758-764, 1946.

[3] N. G. De Bruijn : *Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2n zeros and ones that show each n-letter word exactly once*, TH-Report 75-WSK-06, Technological University Eindhoven, the Netherlands, pp. 114, 1975

[4] H. Bidkhori, S. Kishore: *Counting the spanning trees of a directed line graph*, 2009, `http://arxiv.org/pdf/0910.3442.pdf`

[5] C. Flye Sainte-Marie: *Solution to question nr. 48*, l'Intermédiaire des Mathématiciens, 1894, 107-110.

[6] Pavel A. Pevzner, Tang Haixu, Michael S. Waterman:*An Eulerian path approach to DNA fragment assembly*, Proceedings of the National Academy of Science, vol. 98, Issue 17, p.9748-9753, 2001, `https://dx.doi.org/10.1073/pnas.171285098`

[7] A. de Rivière: *Question nr. 48*, l'Intermédiaire des Mathématiciens, 1894, 19-20.

[8] Daniel R. Zerbino, E. Birney: *Velvet: Algorithms for de novo short read assembly using de Bruijn graphs*, Genome Research 18 (5): 821829, 2008, `https://dx.doi.org/10.1101/gr.074492.107`

[9] R. P. Stanley: *BIJECTIVE PROOF PROBLEMS*, 2009, `http://www-math.mit.edu/~rstan/bij.pdf`, Problem 28.

[10] R. P. Stanley: *TOPICS IN ALGEBRAIC COMBINATORICS*, 2011, `http://www-math.mit.edu/~rstan/algcomb/algcomb.pdf`

Article [[Ru02]]: On the Number of Rich Words

On the Number of Rich Words

Josef Rukavicka(✉)

Department of Mathematics,
Faculty of Nuclear Sciences and Physical Engineering,
Czech Technical University in Prague,
Trojanova 13, 120 01 Prague 2, Czech Republic
josef.rukavicka@seznam.cz

Abstract. Any finite word w of length n contains at most $n+1$ distinct palindromic factors. If the bound $n+1$ is reached, the word w is called rich. The number of rich words of length n over an alphabet of cardinality q is denoted $R_q(n)$. For binary alphabet, Rubinchik and Shur deduced that $R_2(n) \leq c1.605^n$ for some constant c. In addition, Guo, Shallit and Shur conjectured that the number of rich words grows slightly slower than $n^{\sqrt{n}}$. We prove that $\lim_{n\to\infty} \sqrt[n]{R_q(n)} = 1$ for any q, i.e. $R_q(n)$ has a subexponential growth on any alphabet.

Keywords: Rich words · Enumeration · Palindromes · Palindromic factorization

1 Introduction

The study of palindromes is a frequent topic and many diverse results may be found. In recent years, a number of articles deal with so-called *rich* words, or also words having *palindromic defect* 0. They are words having the maximum number of palindromic factors. As noted by [6], a finite word w contains at most $|w| + 1$ distinct palindromic factors with $|w|$ being the length of w. The rich words are exactly those that attain this bound. It is known that on a binary alphabet the set of rich words contains factors of Sturmian words, factors of complementary symmetric Rote words, factors of the period-doubling word, etc., see [1,4,6,13]. On a multiliteral alphabet, the set of rich words contains for example factors of Arnoux—Rauzy words and factors of words coding symmetric interval exchanges.

Rich words can be characterized using various properties, see for instance [2,5,8]. The concept of rich words can also be generalized to respect so-called pseudopalindromes, see [10]. In this paper we focus on an unsolved question of computing the number of rich words of length n over an alphabet with $q > 1$ letters. This number is denoted $R_q(n)$.

This question is investigated in [15], where J. Vesti gives a recursive lower bound on the number of rich words of length n, and an upper bound on the number of binary rich words. Both these estimates seem to be very rough. In [9],

É. Charlier et al. (Eds.): DLT 2017, LNCS 10396, pp. 345–352, 2017.
DOI: 10.1007/978-3-319-62809-7_26

C. Guo, J. Shallit and A.M. Shur construct for each n a large set of rich words of length n. Their construction gives, currently, the best lower bound on the number of binary rich words, namely $R_2(n) \geq \frac{C^{\sqrt{n}}}{p(n)}$, where $p(n)$ is a polynomial and the constant $C \approx 37$. On the other hand, the best known upper bound is exponential. As mentioned in [9], a calculation performed recently by M. Rubinchik provides the upper bound $R_2(n) \leq c1.605^n$ for some constant c, see [11].

Our main result stated as Theorem 10 shows that $R_q(n)$ has a subexponential growth on any alphabet. More precisely, we prove that

$$\lim_{n\to\infty} \sqrt[n]{R_q(n)} = 1\,.$$

In [14], Shur calls languages with the above property *small*. Our result is an argument in favor of a conjecture formulated in [9] saying that for some infinitely growing function $g(n)$ the following holds true $R_2(n) = \mathcal{O}\left(\frac{n}{g(n)}\right)^{\sqrt{n}}$.

To derive our result we consider a specific factorization of a rich word into distinct rich palindromes, here called UPS-factorization (Unioccurrent Palindromic Suffix factorization), see Definition 2. Let us mention that another palindromic factorizations have already been studied, see [3,7]: *Minimal* (minimal number of palindromes), *maximal* (every palindrome cannot be extended on the given position) and *diverse* (all palindromes are distinct). Note that only the *minimal* palindromic factorization has to exist for every word.

The article is organized as follows: Sect. 2 recalls notation and known results. In Sect. 3 we study a relevant property of UPS-factorization. The last section is devoted to the proof of our main result.

2 Preliminaries

Let us start with a couple of definitions: Let A be an alphabet of q letters, where $q > 1$ and $q \in \mathbb{N}$ ($\mathbb{N}$ denotes the set of nonnegative integers). A finite sequence $u_1u_2\cdots u_n$ with $u_i \in A$ is a *finite word*. Its length is n and is denoted $|u_1u_2\cdots u_n| = n$. Let A^n denote the set of words of length n. We define that A^0 contains just the empty word. It is clear that the size of A^n is equal to q^n.
Given $u = u_1u_2\cdots u_n \in A^n$ and $v = v_1v_2\cdots v_k \in A^k$ with $0 \leq k \leq n$, we say that v is a *factor* of u if there exists i such that $0 \leq i$, $i+k \leq n$ and $u_{i+1} = v_1$, $u_{i+2} = v_2$, $\ldots$, $u_{i+k} = v_k$.

A word $u = u_1u_2\cdots u_n$ is called a *palindrome* if $u_1u_2\cdots u_n = u_nu_{n-1}\cdots u_1$. The empty word is considered to be a palindrome and a factor of any word.

A word u of length n is called *rich* if u has $n+1$ distinct palindromic factors. Clearly, $u = u_1u_2\cdots u_n$ is rich if and only if its *reversal* $u_nu_{n-1}\cdots u_1$ is rich as well.

Any factor of a rich word is rich as well, see [8]. In other words, the language of rich words is factorial. In particular it means that $R_q(n)R_q(m) \leq R_q(n+m)$ for any $m, n, q \in \mathbb{N}$. Therefore, the Fekete's lemma implies existence of the limit of $\sqrt[n]{R_q(n)}$ and moreover

$$\lim_{n\to\infty} \sqrt[n]{R_q(n)} = \inf\left\{ \sqrt[n]{R_q(n)} \colon n \in \mathbb{N} \right\}.$$

For a fixed n_0, one can find the number of all rich words of length n_0 and obtain an upper bound on the limit. Using a computer Rubinchik counted $R_2(n)$ for $n \leq 60$, (see the sequence A216264 in OEIS). As $\sqrt[60]{R_2(60)} < 1.605$, he obtained the upper bound given in Introduction.

As shown in [8], any rich word u over an alphabet A is richly prolongable, i.e., there exist letters $a, b \in A$ such that aub is also rich. Thus a rich word is a factor of an arbitrarily long rich word. But the question whether two rich words can appear simultaneously as factors of a longer rich word may have a negative answer. It means that the language of rich words is not recurrent. This fact makes the enumeration of rich words hard.

3 Factorization of Rich Words into Rich Palindromes

Let us recall one important property of rich words [6, Definition 4 and Proposition 3]: The longest palindromic suffix of a rich word w has exactly one occurrence in w (we say that the longest palindromic suffix of w is *unioccurrent* in w). It implies that $w = w^{(1)}w_1$, where w_1 is a palindrome which is not a factor of $w^{(1)}$. Since every factor of a rich word is a rich word as well, it follows that $w^{(1)}$ is a rich word and thus $w^{(1)} = w^{(2)}w_2$, where w_2 is a palindrome which is not a factor of $w^{(2)}$. Obviously $w_1 \neq w_2$. We can repeat the process until $w^{(p)}$ is the empty word for some $p \in \mathbb{N}$, $p \geq 1$. We express these ideas by the following lemma:

Lemma 1. *Let w be a rich word. There exist distinct non-empty palindromes $w_1, w_2, \ldots, w_p$ such that*

$$w = w_p w_{p-1} \cdots w_2 w_1 \text{ and } w_i \text{ is the longest palindromic suffix of } w_p w_{p-1} \cdots w_i \text{ for } i = 1, 2, \ldots, p. \quad (1)$$

Definition 2. *We define UPS-factorization (Unioccurrent Palindromic Suffix factorization) to be the factorization of a rich word w into the form (1).*

Since the w_i in the factorization (1) are non-empty, it is clear that $p \leq n = |w|$. From the fact that the palindromes w_i in the factorization (1) are distinct we can derive a better upper bound on p. The aim of this section is to prove the following theorem:

Theorem 3. *There is a constant $c > 1$ such that for any rich word w of length n the number p of palindromes in the UPS-factorization of $w = w_p w_{p-1} \cdots w_2 w_1$ satisfies*

$$p \leq c\frac{n}{\ln n}. \quad (2)$$

Before proving the theorem, we show two auxiliary lemmas:

Lemma 4. *Let $q, n, t \in \mathbb{N}$ such that*

$$\sum_{i=1}^{t} iq^{\lceil \frac{i}{2} \rceil} \geq n. \tag{3}$$

The number p of palindromes in the UPS-factorization $w = w_p w_{p-1} \cdots w_2 w_1$ of any rich word w with $n = |w|$ satisfies

$$p \leq \sum_{i=1}^{t} q^{\lceil \frac{i}{2} \rceil}. \tag{4}$$

Proof. Let $f_1, f_2, f_3, \ldots$ be an infinite sequence of all non-empty palindromes over an alphabet A with $q = |A|$ letters, where the palindromes are ordered in such a way that $i < j$ implies that $|f_i| \leq |f_j|$. Therefore, the palindromes $f_1, \ldots, f_q$ are of length 1, the palindromes $f_{q+1}, \ldots, f_{2q}$ are of length 2, etc. Since $w_1, \ldots, w_p$ are distinct non-empty palindromes we have $\sum_{i=1}^{p} |f_i| \leq \sum_{i=1}^{p} |w_i| = n$. The number of palindromes of length i over the alphabet A with q letters is equal to $q^{\lceil \frac{i}{2} \rceil}$ (just consider that the "first half" of a palindrome determines its second half). The number $\sum_{i=1}^{t} iq^{\lceil \frac{i}{2} \rceil}$ equals the length of a word obtained as concatenation of all palindromes of length less than or equal to t. Since $\sum_{i=1}^{p} |f_i| \leq n \leq \sum_{i=1}^{t} iq^{\lceil \frac{i}{2} \rceil}$, it follows that the number of palindromes p is less than or equal to the number of all palindromes of length at most t; this explains the inequality (4).

Lemma 5. *Let $N \in \mathbb{N}$, $x \in \mathbb{R}$, $x > 1$ such that $N(x-1) \geq 2$. We have*

$$\frac{Nx^N}{2(x-1)} \leq \sum_{i=1}^{N} ix^{i-1} \leq \frac{Nx^N}{(x-1)}. \tag{5}$$

Proof. The sum of the first N terms of a geometric series with the quotient x is equal to $\sum_{i=1}^{N} x^i = \frac{x^{N+1}-x}{x-1}$. Taking the derivative of this formula with respect to x with $x > 1$ we obtain: $\sum_{i=1}^{N} ix^{i-1} = \frac{x^N(N(x-1)-1)+1}{(x-1)^2} = \frac{Nx^N}{x-1} + \frac{1-x^N}{(x-1)^2}$. It follows that the right inequality of (5) holds for all $N \in \mathbb{N}$ and $x > 1$. The condition $N(x-1) \geq 2$ implies that $\frac{1}{2}N(x-1) \leq N(x-1) - 1$, which explains the left inequality of (5).

We can start the proof of Theorem 3:

Proof (Proof of Theorem 3). Let $t \in \mathbb{N}$ be a minimal nonnegative integer such that the inequality (3) in Lemma 4 holds. It means that:

$$n > \sum_{i=1}^{t-1} iq^{\lceil \frac{i}{2} \rceil} \geq \sum_{i=1}^{t-1} iq^{\frac{i}{2}} = q^{\frac{1}{2}} \sum_{i=1}^{t-1} iq^{\frac{i-1}{2}} \geq \frac{(t-1)q^{\frac{t}{2}}}{2(q^{\frac{1}{2}}-1)}, \tag{6}$$

josef.rukavicka@seznam.cz

where for the last inequality we exploited (5) with $N = t - 1$ and $x = q^{\frac{1}{2}}$. If $q \geq 9$, then the condition $N(x-1) = (t-1)(q^{\frac{1}{2}} - 1) \geq 2$ is fulfilled (it is the condition from Lemma 5) for any $t \geq 2$. Hence let us suppose that $q \geq 9$ and $t \geq 2$. From (6) we obtain:

$$\frac{q^{\frac{t}{2}}}{q^{\frac{1}{2}} - 1} \leq \frac{2n}{t-1} \leq \frac{4n}{t}. \tag{7}$$

Since t is such that the inequality (3) holds and $i \leq q^{\frac{i+1}{2}}$ for any $i \in \mathbb{N}$ and $q \geq 2$, we can write:

$$n \leq \sum_{i=1}^{t} i q^{\frac{i+1}{2}} \leq \sum_{i=1}^{t} q^{i+1} = q^2 \frac{q^t - 1}{q - 1} \leq \frac{q^2}{q-1} q^t \leq q^{2t}. \tag{8}$$

We apply the logarithm on the previous inequality:

$$\ln n \leq 2t \ln q. \tag{9}$$

An upper bound on the number of palindromes p in UPS-factorization follows from (4), (7), and (9):

$$p \leq \sum_{i=1}^{t} q^{\lceil \frac{i}{2} \rceil} \leq \sum_{i=1}^{t} q^{\frac{i+1}{2}} \leq q^{\frac{3}{2}} \frac{q^{\frac{t}{2}}}{q^{\frac{1}{2}} - 1} \leq q^{\frac{3}{2}} \frac{4n}{t} \leq q^{\frac{3}{2}} 8 \ln q \frac{n}{\ln n}. \tag{10}$$

The previous inequality requires that $q \geq 9$ and $t \geq 2$. If $t = 1$ then we can easily derive from (3) that $n \leq q$ and consequently $p \leq n \leq q$. Thus the inequality $p \leq q^{\frac{3}{2}} 8 \ln q \frac{n}{\ln n}$ holds as well for this case. Since every rich word over an alphabet with the cardinality $q < 9$ is also a rich word over the alphabet with the cardinality 9, the estimate (2) in Theorem 3 holds if we set the constant c as follows: $c = \max\{8q^{\frac{3}{2}} \ln q, 8 \cdot 9^{\frac{3}{2}} \ln 9\}$.

Remark 6. Note that in [12] it is shown that most of palindromic factors of a random word of length n are of length close to $\ln(n)$ (compare to Theorem 3).

4 Rich Words Form a Small Language

Recall the definition of a *small* language; the aim of this section is to show that the set of rich words forms a small language, see Theorem 10.

We present a recurrent inequality for $R_q(n)$. To ease our notation we omit the specification of the cardinality of alphabet and write $R(n)$ instead of $R_q(n)$.

Let us define

$$\kappa_n = \left\lceil c \frac{n}{\ln n} \right\rceil,$$

where c is the constant from Theorem 3 and $n \geq 2$.

Theorem 7. *If $n \geq 2$, then*

$$R(n) \leq \sum_{p=1}^{\kappa_n} \sum_{\substack{n_1,n_2,\dots,n_p \geq 1 \\ n_1+n_2+\dots+n_p=n}} R\left(\left\lceil \frac{n_1}{2} \right\rceil\right) R\left(\left\lceil \frac{n_2}{2} \right\rceil\right) \dots R\left(\left\lceil \frac{n_p}{2} \right\rceil\right). \qquad (11)$$

Proof. Given $p, n_1, n_2, \dots, n_p$, let $R(n_1, n_2, \dots, n_p)$ denote the number of rich words with UPS-factorization $w = w_p w_{p-1} \dots w_1$, where $|w_i| = n_i$ for $i = 1, 2, \dots, p$. Note that any palindrome w_i is uniquely determined by its prefix of length $\lceil \frac{n_i}{2} \rceil$; obviously this prefix is rich. Hence the number of words that appear in the UPS-factorization as w_i cannot be larger than $R(\lceil \frac{n_i}{2} \rceil)$. It follows that $R(n_1, n_2, \dots, n_p) \leq R(\lceil \frac{n_1}{2} \rceil) R(\lceil \frac{n_2}{2} \rceil) \dots R(\lceil \frac{n_p}{2} \rceil)$. The sum of this result over all possible p (see Theorem 3) and $n_1, n_2, \dots, n_p$ completes the proof.

Proposition 8. *Let $h > 1$, $K \geq 1$ and $\beta_n = \Theta\left(\frac{n}{\ln n}\right)$ If $\Gamma(n)$ is a sequence of positive integers such that $\Gamma(n) \leq K^{\beta_n} h^{\frac{n+\beta_n}{2}} \left(\frac{\mathrm{e}n}{\beta_n}\right)^{\beta_n}$, then $\lim_{n\to\infty} \sqrt[n]{\Gamma(n)} \leq \sqrt{h}$.*

Proof. For any constant α we have $\lim_{n\to\infty} \alpha^{\frac{\beta_n}{n}} = 1$. Moreover, $\lim_{n\to\infty} \left(\frac{n}{\beta_n}\right)^{\frac{\beta_n}{n}} = 1$. Let us suppose that $\Gamma(n) = K^{\beta_n} h^{\frac{n+\beta_n}{2}} \left(\frac{\mathrm{e}n}{\beta_n}\right)^{\beta_n}$. Using these two equalities we obtain $\lim_{n\to\infty} K^{\frac{\beta_n}{n}} h^{\frac{n+\beta_n}{2n}} \left(\frac{\mathrm{e}n}{\beta_n}\right)^{\frac{\beta_n}{n}} = \lim_{n\to\infty} h^{\frac{1}{2}} h^{\frac{\beta_n}{2n}} = \sqrt{h}$. Since $\sqrt[n]{\Gamma(n)} \leq K^{\frac{\beta_n}{n}} h^{\frac{n+\beta_n}{2n}} \left(\frac{\mathrm{e}n}{\beta_n}\right)^{\frac{\beta_n}{n}}$, we conclude that $\lim_{n\to\infty} \sqrt[n]{\Gamma(n)} \leq \sqrt{h}$.

Next, we show that $R(n)$ satisfies the conditions of Proposition 8 with $\beta_n = \kappa_n$.

Proposition 9. *If $h > 1$ and $K \geq 1$, then $R(n) \leq K^{\kappa_n} h^{\frac{n+\kappa_n}{2}} \left(\frac{\mathrm{e}n}{\kappa_n}\right)^{\kappa_n}$.*

Proof. For any integers $p, n_1, \dots, n_p \geq 1$, the assumption implies that $R(\lceil \frac{n_1}{2} \rceil)$ $R(\lceil \frac{n_2}{2} \rceil) \cdots R(\lceil \frac{n_p}{2} \rceil) \leq K^p h^{\frac{n_1+1}{2}} h^{\frac{n_2+1}{2}} \cdots h^{\frac{n_p+1}{2}} \leq K^p h^{\frac{n+p}{2}}$. Using (11) we obtain:

$$R(n) \leq K^{\kappa_n} h^{\frac{n+\kappa_n}{2}} \sum_{p=1}^{\kappa_n} \sum_{\substack{n_1,n_2,\dots,n_p \geq 1 \\ n_1+n_2+\dots+n_p=n}} 1. \qquad (12)$$

The sum

$$S_n = \sum_{\substack{n_1+n_2+\dots+n_p=n \\ n_1,n_2,\dots,n_p \geq 1}} 1$$

can be interpreted as the number of ways how to distribute n coins between p people in such a way that everyone has at least one coin. That is why $S_n = \binom{n-1}{p-1}$.

It is known (see Appendix for a proof) that

$$\sum_{i=0}^{L} \binom{N}{i} \leq \left(\frac{\mathrm{e}N}{L}\right)^L, \text{ for any } L, N \in \mathbb{Z}^+ \text{ and } L \leq N. \qquad (13)$$

From (12) we can write: $R(n) \leq K^{\kappa_n} h^{\frac{n+\kappa_n}{2}} \left(\frac{\mathrm{e}n}{\kappa_n}\right)^{\kappa_n}$.

josef.rukavicka@seznam.cz

The main theorem of this article is a simple consequence of the previous proposition.

Theorem 10. *Let $R(n)$ denote the number of rich words of length n over an alphabet with q letters. We have $\lim\limits_{n\to\infty} \sqrt[n]{R(n)} = 1$.*

Proof. Let us suppose that $\lim_{n\to\infty} \sqrt[n]{R(n)} = \lambda > 1$. Let $\epsilon > 0$ be such that $\lambda+\epsilon < \lambda^2$. The definition of a limit implies that there is n_0 such that $\sqrt[n]{R(n)} < \lambda+\epsilon$ for any $n > n_0$, i.e. $R(n) < (\lambda+\epsilon)^n$. Let $K = \max\{R(1), R(2), \ldots, R(n_0)\}$. It holds for any $n \in \mathbb{N}$ that $R(n) \leq K(\lambda+\epsilon)^n$. Using Propositions 8 and 9 we obtain $\lim\limits_{n\to\infty} \sqrt[n]{R(n)} \leq \sqrt{\lambda+\epsilon} < \lambda$, and this is a contradiction to our assumption that $\lim\limits_{n\to\infty} \sqrt[n]{R(n)} = \lambda > 1$, it follows that $\lambda = 1$ (obviously $\lambda \geq 1$ since it holds that $R(n+1) \geq R(n) \geq 1$ for all $n > 0$).

Acknowledgments. The author wishes to thank Edita Pelantová and Štěpán Starosta for their useful comments. The author acknowledges support by the Czech Science Foundation grant GAČR 13-03538S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/205/OHK4/3T/14.

Appendix

For the reader's convenience, we provide a proof of the well-known inequality we used in the proof of Proposition 9.

Lemma 11. $\sum_{k=0}^{L} \binom{N}{k} \leq \left(\frac{\mathrm{e}N}{L}\right)^L$, *where $L \leq N$ and $L, N \in \mathbb{Z}^+$ ($\mathbb{Z}^+$ denotes the set of positive integers).*

Proof. Consider $x \in (0,1]$. The binomial theorem states that

$$(1+x)^N = \sum_{k=0}^{N} \binom{N}{k} x^k \geq \sum_{k=0}^{L} \binom{N}{k} x^k.$$

By dividing by the factor x^L we obtain

$$\sum_{k=0}^{L} \binom{N}{k} x^{k-L} \leq \frac{(1+x)^N}{x^L}.$$

Since $x \in (0,1]$ and $k - L \leq 0$, then $x^{k-L} \geq 1$, it follows that

$$\sum_{k=0}^{L} \binom{N}{k} \leq \frac{(1+x)^N}{x^L}.$$

Let us substitute $x = \frac{L}{N} \in (0,1]$ and let us use the inequality $1 + x < \mathrm{e}^x$, that holds for all $x > 0$:

$$\frac{(1+x)^N}{x^L} \leq \frac{\mathrm{e}^{xN}}{x^L} = \frac{\mathrm{e}^{\frac{L}{N}N}}{(\frac{L}{N})^L} = \left(\frac{\mathrm{e}N}{L}\right)^L.$$

josef.rukavicka@seznam.cz

References

1. Balková, L.: Beta-integers and Quasicrystals, Ph. D. thesis, Czech Technical University in Prague and Université Paris Diderot-Paris 7 (2008)
2. Balková, L., Pelantová, E., Starosta, Š.: Sturmian jungle (or garden?) on multiliteral alphabets. RAIRO Theor. Inf. Appl. **44**, 443–470 (2010)
3. Bannai, H., Gagie, T., Inenaga, S., Kärkkäinen, J., Kempa, D., Piątkowski, M., Puglisi, S.J., Sugimoto, S.: Diverse palindromic factorization is NP-complete. In: Potapov, I. (ed.) DLT 2015. LNCS, vol. 9168, pp. 85–96. Springer, Cham (2015). doi:10.1007/978-3-319-21500-6_6
4. Massé, A.B., Brlek, S., Labbé, S., Vuillon, L.: Palindromic complexity of codings of rotations. Theor. Comput. Sci. **412**, 6455–6463 (2011)
5. Bucci, M., De Luca, A., Glen, A., Zamboni, L.Q.: A new characteristic property of rich words. Theor. Comput. Sci. **410**, 2860–2863 (2009)
6. Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy. Theor. Comput. Sci. **255**, 539–553 (2001)
7. Frid, A., Puzynina, S., Zamboni, L.: On palindromic factorization of words. Adv. Appl. Math. **50**, 737–748 (2013)
8. Glen, A., Justin, J., Widmer, S., Zamboni, L.Q.: Palindromic richness. Eur. J. Comb. **30**, 510–531 (2009)
9. Guo, C., Shallit, J., Shur, A.M.: Palindromic rich words and run-length encodings. Inform. Process. Lett. **116**, 735–738 (2016)
10. Pelantová, E., Starosta, Š.: Palindromic richness for languages invariant under more symmetries. Theor. Comput. Sci. **518**, 42–63 (2014)
11. Rubinchik, M., Shur, A.M.: EERTREE: an efficient data structure for processing palindromes in strings. In: Lipták, Z., Smyth, W.F. (eds.) IWOCA 2015. LNCS, vol. 9538, pp. 321–333. Springer, Cham (2016). doi:10.1007/978-3-319-29516-9_27
12. Rubinchik, M., Shur, A.M.: The number of distinct subpalindromes in random words. Fund. Inf. **145**, 371–384 (2016)
13. Schaeffer, L., Shallit, J.: Closed, palindromic, rich, privileged, trapezoidal, and balanced words in automatic sequences. Electr. J. Comb. **23**, P1.25 (2016)
14. Shur, A.M.: Growth properties of power-free languages. Comput. Sci. Rev. **6**, 187–208 (2012)
15. Vesti, J.: Extensions of rich words. Theor. Comput. Sci. **548**, 14–24 (2014)

Article [[Ru03]]: Upper Bound for Palindromic and Factor Complexity of Rich Words

Theoretical Informatics and Applications Will be set by the publisher
Informatique Théorique et Applications

UPPER BOUND FOR PALINDROMIC AND FACTOR COMPLEXITY OF RICH WORDS *

JOSEF RUKAVICKA[1]

Abstract. A finite word w of length n contains at most $n+1$ distinct palindromic factors. If the bound $n+1$ is attained, the word w is called rich. An infinite word w is called rich if every finite factor of w is rich.

Let w be a word (finite or infinite) over an alphabet with $q > 1$ letters, let $\mathrm{Fac}_w(n)$ be the set of factors of length n of the word w, and let $\mathrm{Pal}_w(n) \subseteq \mathrm{Fac}_w(n)$ be the set of palindromic factors of length n of the word w.

We present several upper bounds for $|\,\mathrm{Fac}_w(n)|$ and $|\,\mathrm{Pal}_w(n)|$, where w is a rich word. Let $\delta = \frac{3}{2(\ln 3 - \ln 2)}$. In particular we show that

$$|\,\mathrm{Fac}_w(n)| \leq (4q^2n)^{\delta \ln 2n + 2}.$$

In 2007, Baláži, Masáková, and Pelantová showed that

$$|\,\mathrm{Pal}_w(n)| + |\,\mathrm{Pal}_w(n+1)| \leq |\,\mathrm{Fac}_w(n+1)| - |\,\mathrm{Fac}_w(n)| + 2,$$

where w is an infinite word whose set of factors is closed under reversal. We prove this inequality for every finite word v with $|v| \geq n+1$ and $\mathrm{Fac}_v(n+1)$ closed under reversal.

1991 Mathematics Subject Classification. 68R15.

1. INTRODUCTION

The field of combinatorics on words includes the study of *palindromes* and *rich* words. In recent years there have appeared several articles concerning this topic [3, 5, 8, 17]. Recall that a palindrome is a word that is equal to its *reversal*, such as "noon" and "level". A word is called rich if it contains the maximal number

* *Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague.*
[1] josef.rukavicka@seznam.cz

of palindromic factors. It is known that a word of length n can contain at most $n+1$ palindromic factors, including the empty word [8]. An infinite word w is rich if every finite factor of w is rich.

Rich words possess various properties; see, for instance, [4, 7, 9]. We will use two of them. The first uses the notion of a *complete return*. Given a word w and a factor r of w, we call the factor r a *complete return* to u in w if r contains exactly two occurrences of u, one as a prefix and one as a suffix. A property of rich words is that all complete returns to any palindromic factor u in w are palindromes [9].

The second property of rich words that we use says that a factor r of a rich word w is uniquely determined by its longest palindromic prefix and its longest palindromic suffix [7]. Some generalizations of this property may be found in [12].

In the current article we present upper bounds for the palindromic and factor complexity of rich words. In other words, this means that we derive upper bounds for the number of palindromes and factors of given length in a rich word w. There are already some related results; see below.

We start with some results that hold for arbitrary (not only rich) words.

Let us define $\mathrm{Fac}_w(n)$ to be the set of factors of length n of the word w, let $\mathrm{Pal}_w(n)$ be the set of palindromic factors of length n of w, and let $\mathrm{Fac}_w = \bigcup_{j\geq 0}\mathrm{Fac}_w(j)$, where w is a finite or infinite word. Let w^R denote the reversal of $w = w_1w_2\cdots w_{n-1}w_n$, where w_i are letters; formally $w^R = w_nw_{n-1}\cdots w_2w_1$. We say that a set S of finite words is *closed under reversal* if $w \in S$ implies that $w^R \in S$.

It is clear that $|\,\mathrm{Pal}_w(n)| \leq |\,\mathrm{Fac}_w(n)|$. Some less obvious inequalities are known. One of the interesting inequalities is the following one [2,4]. If w is an infinite word with Fac_w closed under reversal then

$$|\,\mathrm{Pal}_w(n)| + |\,\mathrm{Pal}(w, n+1)| \leq |\,\mathrm{Fac}_w(n+1)| - |\,\mathrm{Fac}_w(n)| + 2. \tag{1}$$

In [2] the authors proved the inequality (1) for uniformly recurrent words, but in the proof only "recurrent" is applied. It is known that if Fac_w is closed under reversal, then w is recurrent [6, Proposition 2.2]. In Section 3 we generalize (1) for every finite word v with $\mathrm{Fac}_v(n+1)$ closed under reversal, which allows us to improve our upper bound from Section 2 for the factor complexity of finite rich words.

In [1], another inequality has been proven for infinite non-ultimately periodic words: $|\,\mathrm{Pal}_w(n)| < \frac{16}{n}|\,\mathrm{Fac}_w(n + \lfloor\frac{n}{4}\rfloor)|$.

In [14], the authors show that a random word of length n contains, on expectation, $\Theta(\sqrt{n})$ distinct palindromic factors.

Now, let us focus on rich words.

Let $\Pi(n)$ denote the number of rich words of length n. If w is a rich word then obviously $|\,\mathrm{Fac}_w(n)| \leq \Pi(n)$. Hence the number of rich words forms the upper bound for the palindromic and factor complexity of rich words. The number of rich words was investigated in [19], where the author gives a recursive lower bound on the number of rich words of length n, and an upper bound on the number of binary rich words. Better results can be found in [10]. The authors of [10] construct for each n a large set of rich words of length n. Their construction gives, currently, the best lower bound on the number of binary rich words, namely

$$\Pi(n) \geq \frac{C^{\sqrt{n}}}{p(n)}, \tag{2}$$

where $p(n)$ is a polynomial and the constant $C \approx 37$.
Every factor of a rich word is also rich [9]. In other words, the language of rich words is factorial. In particular, this means that $\Pi(n)\Pi(m) \geq \Pi(n+m)$ for all $m, n \in \mathbb{N}$. Therefore, Fekete's lemma implies the existence of the limit of $\sqrt[n]{\Pi(n)}$, and moreover [10]:

$$\lim_{n\to\infty} \sqrt[n]{\Pi(n)} = \inf\left\{ \sqrt[n]{\Pi(n)} \colon n \in \mathbb{N} \right\}.$$

For a fixed n_0, one can find the number of all rich words of length n_0 and obtain an upper bound on the limit. Using a computer Rubinchik counted $\Pi(n)$ for $n \leq 60$; see the sequence `https://oeis.org/A216264`. As $\sqrt[60]{\Pi(60)} < 1.605$, he obtained an upper bound for the binary alphabet: $\Pi(n) < c1.605^n$ for some constant c [10].

In [15], the author shows that $\Pi(n)$ has a subexponential growth on every finite alphabet. Formally $\lim\limits_{n\to\infty} \sqrt[n]{\Pi(n)} = 1$. This result is an argument in favor of a conjecture formulated in [10] saying that for some infinitely growing function $g(n)$ the following holds for a binary alphabet:

$$\Pi(n) = \mathcal{O}\Big(\frac{n}{g(n)}\Big)^{\sqrt{n}}.$$

As already mentioned, we construct upper bounds for palindromic and factor complexity of rich words. The proof uses the following idea. Let u be a palindromic factor of a rich word w on the alphabet A, such that aub is factor of w, where $a, b \in A$ and $a \neq b$. Let $\mathrm{lpp}(w)$ and $\mathrm{lps}(w)$ denote the longest palindromic prefix and suffix of w respectively. Then $\mathrm{lpp}(aub)$ and $\mathrm{lps}(aub)$ uniquely determine the factor aub in w [7]. Let $\mathrm{lpps}(w)$ denote the longest proper palindromic suffix of w. We show that a, b and $\mathrm{lpps}(u)$ also uniquely determine aub. In addition, we observe that either $|\,\mathrm{lpps}(u)| \leq \frac{1}{2}|u|$ or u contains a palindromic factor $\bar{u}$ that uniquely determines u and $|\bar{u}| \leq \frac{1}{2}|u|$. We obtain a "short" palindrome and letters a, b that uniquely determine the "long" palindrome u in the case when aub is a factor of w. In these "short" palindromes there are again other "shorter" palindromes, and so on. As a consequence we present an upper bound for the number of factors of the form aub with $|aub| = n$. The property of rich words that all complete returns to any palindromic factor u in w are palindromes [9] allows us to prove that if w contains the factors xux and yuy, where $x, y \in A$ and $x \neq y$, then w must contain a factor of the form aub, where $a, b \in A$ and $a \neq b$. This property demonstrates the relation between the factors aub and palindromic factors xux. Due to this we derive an upper bound for the palindromic complexity of rich words. With the upper bound for palindromic complexity, the property that each factor is uniquely determined by its longest palindromic prefix and suffix [7], and the inequality (1) we obtain several upper bounds on palindromic and factor complexity. The main result of the current article is the following theorem.

Theorem 1.1. *If $\delta = \frac{3}{2(\ln 3 - \ln 2)}$, w is a finite or infinite rich word over an alphabet with $q > 1$ letters, and n is a positive integer then*

$$|\operatorname{Fac}_w(n)| \leq (4q^2n)^{\delta \ln 2n+2}.$$

The main result is a quasi-polynomial upper bound for factor complexity of rich words. This is much less than the lower bound on the number of rich words; recall (2). Thus an infinite rich word can contain only a small share of all finite rich words. This contrasts with power-free languages, where an infinite word can contain all extendable finite words with the same power-freeness restriction [13, 16, 18].

2. Palindromic and factor complexity of rich words

Consider an alphabet A with q letters, where $q > 1$. Let $A^+ = \bigcup_{j>0} A^j$ denote the set of all nonempty words over A, where A^j is the set of words of length j.

Let ϵ denote the empty word, let $A^* = A^+ \cup \{\epsilon\}$, and let

$$A^\infty = \{w_1w_2w_3\cdots \mid w_i \in A \text{ and } i > 0\}$$

be the set of infinite words.

Let $R_n \subseteq A^n$ be the set of rich words of length $n \geq 0$. Let $R^+ = \bigcup_{j>0} R_j$ and $R^* = R^+ \cup \{\epsilon\}$. In addition, we define $R^\infty \subseteq A^\infty$ to be the set of infinite rich words. Let $R = R^+ \cup R^\infty$.

Let $\operatorname{lps}(w)$ and $\operatorname{lpp}(w)$ be the longest palindromic suffix and the longest palindromic prefix of a word $w \in A^*$ respectively. Additionally, we introduce $\operatorname{lpps}(w)$ to be the longest proper palindromic suffix and $\operatorname{lppp}(w)$ to be the longest proper palindromic prefix, where $|w| > 1$; proper means that $\operatorname{lpps}(w) \neq w$ and $\operatorname{lppp}(w) \neq w$. For a word w with $|w| \leq 1$ we define $\operatorname{lppp}(w) = \operatorname{lpps}(w) = \epsilon$.

Let $w = w_1w_2\cdots w_n$ be a word, where $w_i \in A$. We define $w[i] = w_i$ and $w[i,j] = w_iw_{i+1}\cdots w_j$, where $0 < i \leq j \leq n$.

Moreover we define the following notation:

- $\mathrm{P}_n \subset A^n$: the set of palindromes of length $n \geq 0$.
- $\mathrm{P}^+ = \bigcup_{j>0} \mathrm{P}_j$ (the set of all nonempty palindromes).
- Fac_w: the set of factors of the word $w \in A^* \cup A^\infty$.
- $\operatorname{Fac}_w(n) = \{u \mid u \in \operatorname{Fac}_w \text{ and } |u| = n\}$ (the set of factors of length n).
- $\operatorname{Pal}_w = (\mathrm{P}^+ \cup \{\epsilon\}) \cap \operatorname{Fac}_w$ (the set of palindromic factors).
- $\operatorname{Pal}_w(n) = \operatorname{Fac}_w(n) \cap \mathrm{P}_n$ (the set of palindromic factors of length n).

Definition 2.1. *Let* $\operatorname{trim}(w) = w[2, |w|-1]$*, where* $w \in A^*$ *and* $|w| > 2$*. For* $|w| \leq 2$ *we define* $\operatorname{trim}(w) = \epsilon$*. If* S *is a set of words, then*

$$\operatorname{trim}(S) = \{\operatorname{trim}(v) \mid v \in S\}.$$

Remark 2.2. The function $\operatorname{trim}(w)$ removes the first and last letter from w.

Example 2.3. Suppose that $A = \{0, 1, 2, 3, 4, 5\}$.

- $\operatorname{trim}(01123501) = 112350$.
- $\operatorname{trim}(\{12213, 112, 2, 344\}) = \{221, 1, \epsilon, 4\}$.

We will deal a lot with the words of the form aub, where u is a palindrome and a, b are distinct letters. Hence we introduce some more notation for them.

Definition 2.4. *Given $w \in R$ and $n > 2$, let*

$$\mathrm{Sw}_w(n) = \{aub \mid aub \in \mathrm{Fac}_w(n) \text{ and } u \in \mathrm{Pal}_w(n-2) \text{ and } a, b \in A \text{ and } a \neq b\}.$$

If $n \leq 2$ then we define $\mathrm{Sw}_w(0) = \mathrm{Sw}_w(1) = \mathrm{Sw}_w(2) = \emptyset$.

*Let $\bar{\mathrm{Sw}}_w(n) = \bigcup_{aub \in \mathrm{Sw}_w(n)} \{(u, a), (u, b)\}$, where $a, b \in A$. Let $aub \in \mathrm{Sw}_w(n)$, where $a, b \in A$. We call the word aub a u-*switch *of w. Alternatively we say that w contains a u-*switch.

Remark 2.5. Note that a pair $(u, a) \in \bar{\mathrm{Sw}}_w(n)$ if and only if there exists $b \in A$ such that $aub \in \mathrm{Sw}_w(n)$ or $bua \in \mathrm{Sw}_w(n)$.

Example 2.6. If $A = \{0, 1, 2, 3, 4, 5, 6\}$ and

$$w = 5112211311001131133114111146$$

then:

- $\mathrm{Sw}_w(8) = \{51122113, 31133114, 14111146\}$.
- $\operatorname{trim}(\mathrm{Sw}_w(8)) = \{112211, 113311, 411114\}$.
- $\bar{\mathrm{Sw}}_w(8) = \{(112211, 3), (112211, 5), (113311, 3), (113311, 4), (411114, 1), (411114, 6)\}$.
- w does not contain 110011-switch. Formally $110011 \notin \operatorname{trim}(\mathrm{Sw}_w(8))$.

Remark 2.7. The idea of a u-switch is inspired by the next lemma. If a rich word w contains palindromes aua, bub, where $a, b \in A$, $a \neq b$, and $|aua| = |bub| = n$, then w contains a u-switch of length n. The u-switch "switches" from a to b. Note that $aua, bub \in \mathrm{Fac}_w$ does not imply that $aub \in \mathrm{Fac}_w$ or $bua \in \mathrm{Fac}_w$. It may be, for example, that $auc, cub \in \mathrm{Fac}_w$. Nonetheless $(u, a), (u, b) \in \bar{\mathrm{Sw}}_w(n)$.

Lemma 2.8. *Suppose $w \in R$ and suppose $u \in \mathrm{Pal}_w(n-2)$, where $n > 2$. If $a, b_1, b_2 \in A$, $|\{a, b_1, b_2\}| > 1$, and $aua, b_1ub_2 \in \mathrm{Fac}_w(n)$ then $(u, a) \in \bar{\mathrm{Sw}}_w(n)$.*

Remark 2.9. The condition $|\{a, b_1, b_2\}| > 1$ in Lemma 2.8 means that at least one letter is different from the others.

Proof. Let r be a factor of w such that aua is unioccurrent in r and $\operatorname{trim}(r)$ is a complete return to u in w. Since aua and b_1ub_2 are factors of w, it is obvious that such r exists. Clearly there are $x_1, x_2, y_1, y_2 \in A$ such that x_1ux_2 is a prefix of r and y_1uy_2 is a suffix of r. The complete return $\operatorname{trim}(r)$ to u is a palindrome [9]. Hence $x_2 = y_1$. Since aua is unioccurrent in r, it follows that $x_2 = y_1 = a$, $x_1 \neq y_2$, and $a \in \{x_1, y_2\}$. In consequence we have that $(u, a) \in \bar{\mathrm{Sw}}_w(n)$. □

To clarify the previous proof, let us consider the following two examples. For both examples suppose that $A = \{1, 2, 3, 4, 5, 6\}$.

Example 2.10. Let $w = 321234321252126$. Let $aua = 32123$ and $b_1ub_2 = 52126$. Then $r = 32123432125$ and $\mathrm{trim}(r) = 212343212$ is a complete return to 212. Therefore $(212, 3) \in \bar{\mathrm{Sw}}_w(5)$. Note that b_1ub_2 is not a factor of r.

Example 2.11. Let $w = 321234321252$. Let $aua = 32123$ and $xuy = b_1ub_2 = 32125$. Then $r = 32123432125$ and $\mathrm{trim}(r) = 212343212$ is a complete return to 212. Therefore $(212, 3) \in \bar{\mathrm{Sw}}_w(5)$. Note that b_1ub_2 is a factor of r.

We show that the number of palindromic factors and the number of u-switches are related.

Proposition 2.12. *If $w \in R$ and $n > 2$ then*

$$2|\,\mathrm{Sw}_w(n)| + |\,\mathrm{Pal}_w(n-2)| \geq |\,\mathrm{Pal}_w(n)|.$$

Proof. Let $\omega(w, n) = \{aua | (u, a) \in \bar{\mathrm{Sw}}_w(n)\}$. Less formally said, $\omega(w, n)$ is a set of palindromes of length n such that if w contains a u-switch aub then $aua, bub \in \omega(w, n)$. Obviously we have that

$$|\omega(w, n)| \leq 2|\,\mathrm{Sw}_w(n)|. \tag{3}$$

Let

$$\tilde{\mathrm{Pal}}_w(n) = \{v \mid v \in \mathrm{Pal}_w(n) \text{ and } \mathrm{trim}(v) \in \mathrm{trim}(\mathrm{Sw}_w(n))\}$$

and

$$\dot{\mathrm{Pal}}_w(n) = \{v \mid v \in \mathrm{Pal}_w(n) \text{ and } \mathrm{trim}(v) \notin \mathrm{trim}(\mathrm{Sw}_w(n))\}.$$

Obviously $\mathrm{Pal}_w(n) = \tilde{\mathrm{Pal}}_w(n) \cup \dot{\mathrm{Pal}}_w(n)$ and $\tilde{\mathrm{Pal}}_w(n) \cap \dot{\mathrm{Pal}}_w(n) = \emptyset$. It follows that

$$|\tilde{\mathrm{Pal}}_w(n)| + |\dot{\mathrm{Pal}}_w(n)| = |\,\mathrm{Pal}_w(n)|. \tag{4}$$

Suppose $v \in \mathrm{Pal}_w(n)$ and let $u = \mathrm{trim}(v)$.

- If $v \in \tilde{\mathrm{Pal}}_w(n)$ then w contains a u-switch. From Lemma 2.8 it follows that $v \in \omega(w, n)$; this and (3) imply that

$$|\tilde{\mathrm{Pal}}_w(n)| \leq |\omega(w, n)| \leq 2|\,\mathrm{Sw}_w(n)|. \tag{5}$$

- If $v \notin \tilde{\mathrm{Pal}}_w(n)$ then w does not contain a u-switch. We have that $u \in \mathrm{Pal}_w(n-2) \setminus \mathrm{trim}(\mathrm{Sw}_w(n))$. Obviously if $t \in \mathrm{Pal}_w(n-2) \setminus \mathrm{trim}(\mathrm{Sw}_w(n))$, $a, b \in A$, and w has palindromic factors ata and btb, then $a = b$ since w does not contain a t-switch. It follows that

$$|\dot{\mathrm{Pal}}_w(n)| \leq |\,\mathrm{Pal}_w(n-2)|. \tag{6}$$

The proposition follows from (4), (5), and (6). $\square$

To clarify the previous proof, let us consider the following example.

Example 2.13. If $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ and

$$w = 211011233321101145411011611011677877 6$$

then

- $\mathrm{Sw}_w(7) = \{2110114, 4110116\}$,
- $\mathrm{Pal}_w(7) = \{1233321, 2110112, 1145411, 6110116, 6778776\}$,
- $\widetilde{\mathrm{Pal}}_w(7) = \{2110112, 6110116\}$,
- $\dot{\mathrm{Pal}}_w(7) = \{1233321, 1145411, 6778776\}$,
- $\mathrm{Pal}_w(5) = \{23332, 11011, 14541, 77877\}$,
- $2|\,\mathrm{Sw}_w(7)| + |\,\mathrm{Pal}_w(5)| \geq |\,\mathrm{Pal}_w(7)|$, and
- $4 + 4 > 5$.

In the next proposition we show that if a, b are different letters and aub is a switch of a rich word w then the longest proper palindromic suffix r of u and the letters a, b uniquely determine the palindromic factor $u \in \mathrm{Pal}_w$.

Proposition 2.14. *If $w \in R$, $u, v \in \mathrm{Pal}_w$, $\mathrm{lpps}(u) = \mathrm{lpps}(v)$, $a, b \in A$, $a \neq b$, and $aub, avb \in \mathrm{Fac}_w$ then $u = v$.*

Proof. It is known that if r, t are two factors of a rich word w and $\mathrm{lps}(r) = \mathrm{lps}(t)$ and $\mathrm{lpp}(r) = \mathrm{lpp}(t)$, then $r = t$ [7]. We will identify a u-switch by the longest proper palindromic suffix of u and two distinct letters a, b instead of by the functions lps and lpp.

Given a u-switch aub where $a \neq b$, $a, b \in A$, we know that $\mathrm{lps}(aub)$ and $\mathrm{lpp}(aub)$ uniquely determine the factor aub in w. We will prove that for given $a, b \in A$, $a \neq b$, $n \geq 0$, and a palindrome r there is at most one palindrome $u \in \mathrm{Pal}_w$ such that $\mathrm{lpps}(u) = r$ and $aub \in \mathrm{Sw}_w(|aub|)$.

Suppose, to get a contradiction, that there are $u, v \in \mathrm{Pal}_w$, $u \neq v$, $a, b \in A$, $a \neq b$ such that $\mathrm{lps}(aub) = bpb$, $\mathrm{lps}(avb) = bsb$, $\mathrm{lpp}(aub) = axa$, $\mathrm{lpp}(avb) = aya$, $\mathrm{lpps}(u) = \mathrm{lpps}(v) = r$, and $aub, avb \in \bigcup_{j>0} \mathrm{Sw}_w(j)$. This implies that p, s, x, y are prefixes of r. Thus if $x \neq y$, then $|x| \neq |y|$. Without loss of generality, let $|x| < |y|$. Since y is a prefix of r, either ya is a prefix of r or $r = y$. Consequently aya is a prefix of both aub and avb, and this contradicts the assumption that $\mathrm{lpp}(aub) = axa$; aya is a prefix of aub and $|aya| > |axa|$. Analogously if $p \neq s$. It follows that $x = y$ and $p = s$. Therefore $\mathrm{lpp}(aub) = \mathrm{lpp}(avb)$ and $\mathrm{lps}(aub) = \mathrm{lps}(avb)$, which would imply that $u = v$, which is a contradiction.

Hence we conclude that $a, b \in A$, $a \neq b$, and a palindrome r determine at most one palindrome $u \in \mathrm{Pal}_w$ such that $\mathrm{lpps}(u) = r$ and $u \in \mathrm{trim}(\mathrm{Sw}_w(|u| + 2))$. $\square$

In the following we derive an upper bound for the number of u-switches. We need one more definition to be able to partition the set $\mathrm{Sw}_w(n)$ into subsets based on the longest proper palindromic suffix.

Definition 2.15. *Given $w \in R$, $r \in R^+$ and $n \geq 0$, let*

$$\Upsilon_w(n, r) = \{u \mid u \in \mathrm{Sw}_w(n) \text{ and } \mathrm{lpps}(\mathrm{trim}(u)) = r\}.$$

Remark 2.16. The set $\Upsilon_w(n,r)$ contains switches avb of length n of the word w such that the longest proper palindromic suffix of v equals to r, where a,b are letters. Obviously $\bigcup_{r\in \mathrm{Pal}_w} \Upsilon_w(n,r) = \mathrm{Sw}_w(n)$ and $\Upsilon_w(n,r)\cap\Upsilon_w(n,\bar{r}) = \emptyset$ if $r\neq\bar{r}$.

A simple corollary of the previous proposition is that the size of the set $\Upsilon_w(n,r)$ is limited by the constant $q(q-1)$. Recall that q is the size of the alphabet A.

Corollary 2.17. *If $w,r\in R$ and $n\geq 0$ then $|\Upsilon_w(n,r)|\leq q(q-1)$.*

Proof. From Proposition 2.14 it follows that

$$|\Upsilon_w(n,r)|\leq |\{(a,b)\mid a,b\in A \text{ and } a\neq b\}| = q(q-1).$$

In other words, $|\Upsilon_w(n,r)|$ is equal or smaller that the number of pairs of distinct letters (a,b). □

We define $\bar{\Gamma}_w(n) = \max\{|\mathrm{Sw}_w(i)| \mid 0\leq i\leq n\}$, where $w\in R$ and $n\geq 0$. Furthermore we define $\Gamma_w(n) = \max\{q, \bar{\Gamma}_w(n)\}$.

Remark 2.18. We defined $\Gamma_w(n)$ as the maximum from the set of sizes of $\mathrm{Sw}_w(i)$, where $0\leq i\leq n$. In addition, we defined that $\Gamma_w(n)\geq q$. This is just for practical reason to make the formulas easier; since we look for upper bounds, this simplification is justified. The function $\Gamma_w(n)$ will allow us to present another relation between the number of palindromic factors of length n and the number of u-switches without using $\mathrm{Pal}_w(n-2)$.

Lemma 2.19. *If $w\in R$ and $n>0$ then*

$$n\Gamma_w(n)\geq |\mathrm{Pal}_w(n)|.$$

Proof. We define two functions $\bar{\phi}$ and ϕ as follows. If n is even then $\bar{\phi}(n)=2$, otherwise $\bar{\phi}(n)=1$. Let $\phi(n)=\{2+\bar{\phi}(n), 4+\bar{\phi}(n),\dots,n\}$. For example $\phi(8)=\{4,6,8\}$ and $\phi(9)=\{3,5,7,9\}$.

Proposition 2.12 states that

$$2|\mathrm{Sw}_w(n)| + |\mathrm{Pal}_w(n-2)|\geq |\mathrm{Pal}_w(n)|. \tag{7}$$

It follows that

$$2|\mathrm{Sw}_w(n-2)| + |\mathrm{Pal}_w(n-4)|\geq |\mathrm{Pal}_w(n-2)|. \tag{8}$$

From (7) and (8):

$$2|\mathrm{Sw}_w(n)| + 2|\mathrm{Sw}_w(n-2)| + \mathrm{Pal}_w(n-4)|\geq |\mathrm{Pal}_w(n)|. \tag{9}$$

In general (7) implies that

$$2|\mathrm{Sw}_w(n-i)| + |\mathrm{Pal}_w(n-2i)|\geq |\mathrm{Pal}_w(n-i)|. \tag{10}$$

Then by iterative applying of (10) to (9) we obtain that

$$\sum_{j\in\phi(n)} 2|\operatorname{Sw}_w(j)| + |\operatorname{Pal}_w(\bar{\phi}(n))| \geq |\operatorname{Pal}_w(n)| \tag{11}$$

We have that $|\operatorname{Pal}_w(\bar{\phi}(n))| \leq q$; just consider that $|\operatorname{Pal}_w(\bar{\phi}(n))|$ is the number of palindromes of length 1 or 2. Recall that $\Gamma_w(n) \geq |\operatorname{Sw}_w(j)|$ for $2 < j < n$ and realize that $|\phi(n)| \leq \frac{n-1}{2}$. It follows from (11) that $(n-1)\Gamma_w(n) + q \geq |\operatorname{Pal}_w(n)|$. It is easy to see that $n\Gamma_w(n) \geq (n-1)\Gamma_w(n) + q$ for $n > 0$, since $\Gamma_w(n) \geq q$. This completes the proof. □

We will need to cope with the longest proper palindromic suffixes that are "too long". We show that if the longest proper palindromic suffix $\operatorname{lpps}(v)$ is longer the half of the length of v, then v contains a "short" palindromic factor, that uniquely determines v. We will use the two following lemmas from [11]:

Lemma 2.20. *(see [11, Lemma 1]) Suppose p is a period of a nonempty palindrome w; then there are palindromes a and b such that $|ab| = p$, $b \neq \epsilon$, and $w = (ab)^j a$ for some non-negative integer j.*

Lemma 2.21. *(see [11, Lemma 2]) Suppose w is a palindrome and u is its proper suffix-palindrome or prefix-palindrome; then the number $|w| - |u|$ is a period of w.*

Let $u, v \in \mathrm{P}^+$ such that u is a suffix of v and $|u| < |v|$. Lemma 2.21 implies that v is periodic with period $p = |v| - |u|$. Lemma 2.20 implies that there are palindromes a, b such that b is nonempty and $p = |ab|$ and $v = (ab)^j a$ for some non-negative integer j. We define $\bar{\rho}(u,v) = (a,b)$ and $\rho(u,v) = aba \in \mathrm{P}^+$.

The next lemma is an obvious consequence of Lemma 2.20 and Lemma 2.21. It says that v is uniquely determined by the palindrome $\rho(u,v)$ and by the lengths of u and v.

Lemma 2.22. *If $u_1, u_2, v_1, v_2 \in \mathrm{P}^+$, $|v_1| = |v_2|$, $|u_1| = |u_2|$, $|u_1| < |v_1|$, u_1 is a suffix of v_1, u_2 is a suffix of v_2, and $\rho(u_1, v_1) = \rho(u_2, v_2)$ then $v_1 = v_2$.*

Proof. Let $\bar{\rho}(u_1, v_1) = (a_1, b_1)$ and let $\rho(u_2, v_2) = (a_2, b_2)$. Let $p = |v_1| - |u_1| = |v_2| - |u_2|$. Since $\rho(u_1, v_1) = \rho(u_2, v_2)$, from Lemma 2.20 and Lemma 2.21 we have that $p = |a_1 b_1| = |a_2 b_2|$. Also it follows that $a_1 b_1 = a_2 b_2$ and $a_1 b_1 a_1 = a_2 b_2 a_2$. In consequence we get that $a_1 = a_2$ and $b_1 = b_2$. This ends the proof. □

In the next lemma we consider a palindromic suffix u of a palindrome v, which is longer than the half of v. For this case we show an upper bound for the length of the palindrome $\rho(u,v)$.

Lemma 2.23. *If $u, v \in \mathrm{P}^+$, u is a suffix of v, and $\frac{1}{2}|v| \leq |u| < |v|$ then*

$$\rho(u,v) \leq \lceil \frac{2}{3}|v| \rceil.$$

Proof. Let $(a,b) = \bar{\rho}(u,v)$. It is easy to verify that $\frac{1}{2}|v| \leq |u| < |v|$ implies that $j \geq 2$, where $v = (ab)^j a$.

Let c be a positive real constant such that $|aba| = c|(ab)^j a|$. For given a, b it is clear that c decreases as j increases. Since $j > 1$ it follows that c is maximal for $j = 2$. Thus $c \leq \frac{|aba|}{|ababa|} = \frac{2|a|+|b|}{3|a|+2|b|}$. The lemma follows. □

We derive an upper bound for the number of u-switches.

Proposition 2.24. *If $w \in R$ and $n > 2$ then*

$$\Gamma_w(n) \leq 2q^2(\frac{2n}{3})^3 \Gamma_w(\lfloor \frac{2n}{3} \rfloor).$$

Proof. We partition the set $\mathrm{Sw}_w(n)$ into sets $\Delta_\rho(w,n), \Delta_{\mathrm{lpps}}(w,n)$ as follows. Let $avb \in \mathrm{Sw}_w(n)$ be a v-switch, where $a, b \in A$. If $\frac{1}{2}|v| \leq |\,\mathrm{lpps}(v)|$ then $avb \in \Delta_\rho(w,n)$ otherwise $avb \in \Delta_{\mathrm{lpps}}(w,n)$. Obviously $\Delta_\rho(w,n) \cap \Delta_{\mathrm{lpps}}(w,n) = \emptyset$ and

$$\mathrm{Sw}_w(n) = \Delta_\rho(w,n) \cup \Delta_{\mathrm{lpps}}(w,n). \tag{12}$$

Let us investigate the sizes of $\Delta_\rho(w,n)$ and $\Delta_{\mathrm{lpps}}(w,n)$.

- If $avb \in \Delta_\rho(w,n)$ then let $u = \mathrm{lpps}(v)$. We have that $\rho(u,v)$, $|u|$, and $|v|$ uniquely determine the palindrome v; see Lemma 2.22. In addition, $|\rho(u,v)| \leq \lceil \frac{2|v|}{3} \rceil$; see Lemma 2.23. Realize that $|v| = n - 2$; then the number of all palindromic factors of w of length $\leq \lceil \frac{2(n-2)}{3} \rceil$ multiplied by $\lceil \frac{n-2}{2} \rceil$ (the number of different values of $|u|$) must be bigger or equal to the size of $\mathrm{trim}(\Delta_\rho(w,n))$. Realize that the set $\mathrm{trim}(\Delta_\rho(w,n))$ contains palindromes of length $n-2$. Since $\lceil \frac{2(n-2)}{3} \rceil \leq \lfloor \frac{2n}{3} \rfloor$ we have that

$$|\,\mathrm{trim}(\Delta_\rho(w,n))| \leq \lceil \frac{n-2}{2} \rceil \sum_{j=1}^{\lfloor \frac{2n}{3} \rfloor} |\,\mathrm{Pal}_w(j)|. \tag{13}$$

  Since a, b are distinct letters it follows that

$$|\Delta_\rho(w,n))| \leq q(q-1)|\,\mathrm{trim}(\Delta_\rho(w,n))|. \tag{14}$$

- If $avb \in \Delta_{\mathrm{lpps}}(w,n)$ then $|\,\mathrm{lpps}(v)| < \frac{1}{2}|v| = \frac{n-2}{2}$. Obviously we have that

$$\Delta_{\mathrm{lpps}}(w,n) = \bigcup_{r \in S} \Upsilon_w(n,r), \text{ where } S = \{r \mid r \in \mathrm{Pal}_w \text{ and } |r| < \frac{n-2}{2}\}. \tag{15}$$

  Since $\lceil \frac{n-2}{2} \rceil \leq \lfloor \frac{n}{2} \rfloor$ we have from Corollary 2.17 and (15) that

$$|\Delta_{\mathrm{lpps}}| \leq q(q-1) \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} |\,\mathrm{Pal}_w(j)|. \tag{16}$$

It follows from (12), (13), (14), and (16) that

$$|\operatorname{Sw}_w(n)| \leq 2q(q-1)\lceil\frac{n-2}{2}\rceil \sum_{j=1}^{\lfloor\frac{2n}{3}\rfloor} |\operatorname{Pal}_w(j)|. \tag{17}$$

From Lemma 2.19 we know that $|\operatorname{Pal}_w(j)| \leq j\Gamma_w(j)$. Therefore we have that

$$\sum_{j=1}^{\lfloor\frac{2n}{3}\rfloor} |\operatorname{Pal}_w(j)| \leq \sum_{j=1}^{\lfloor\frac{2n}{3}\rfloor} j\Gamma_w(j) \leq \frac{2n}{3}\frac{2n}{3}\Gamma_w(\lfloor\frac{2n}{3}\rfloor). \tag{18}$$

To simplify the formulas, we apply that $q(q-1) < q^2$ and that $\lceil\frac{n-2}{2}\rceil \leq \frac{2n}{3}$. From (17) and (18):

$$|\operatorname{Sw}_w(n)| \leq 2q^2(\frac{2n}{3})^3\Gamma_w(\lfloor\frac{2n}{3}\rfloor). \tag{19}$$

From Definition of $\Gamma_w(n)$ and (19) we get that

$$\Gamma_w(n) = \max\{q, \max\{|\operatorname{Sw}_w(j)| \mid 0 \leq j \leq n\}\} \leq 2q^2(\frac{2n}{3})^3\Gamma_w(\lfloor\frac{2n}{3}\rfloor).$$

This ends the proof. □

We will need the following lemma in the proof of Corollary 2.26.

Lemma 2.25. *If $\beta > 1$ is a real constant then $\prod_{j\geq 1}^{k} \frac{n}{\beta^j} \leq n^{\frac{\ln n}{2\ln\beta}}$, where $k = \lfloor\frac{\ln n}{\ln\beta}\rfloor$.*

Proof.

$$\prod_{j\geq 1}^{k} \frac{n}{\beta^j} = \frac{n}{\beta}\frac{n}{\beta^2}\frac{n}{\beta^3}\cdots\frac{n}{\beta^{k-1}}\frac{n}{\beta^k} \leq \frac{n^k}{\prod_{j=1}^{k}\beta^j}. \tag{20}$$

We have that

$$\prod_{j=1}^{k}\beta^j = \beta\beta^2\beta^3\cdots\beta^{k-1}\beta^k = \beta^{\sum_{j=1}^{k} j} = \beta^{\frac{k(k+1)}{2}}. \tag{21}$$

Then from (20) and (21): $\prod_{j\geq 1}^{k}\frac{n}{\beta^j} \leq \frac{n^k}{\beta^{\frac{k(k+1)}{2}}} = \left(\frac{n}{\beta^{\frac{(k+1)}{2}}}\right)^k$.

Since $\beta^{k+1} \geq n$:

$\left(\frac{n}{\beta^{\frac{(k+1)}{2}}}\right)^k \leq \left(\frac{n}{n^{\frac{1}{2}}}\right)^k = (n^{\frac{1}{2}})^k \leq n^{\frac{\ln n}{2\ln\beta}}$. This completes the proof. □

In order to simplify the notation let $\alpha = \frac{3}{2}$ and let $\delta = \frac{3}{2\ln\alpha} = \frac{3}{2(\ln 3-\ln 2)}$. Based on Proposition 2.24 we will derive a non-recurrent upper bound for $\Gamma_w(n)$.

Corollary 2.26. *If $w \in R$, and $n > 2$ then*

$$\Gamma_w(n) \leq q(2q^2n)^{\delta\ln n}.$$

Proof. Proposition 2.24 states that

$$\Gamma_w(n) \le 2q^2(\frac{n}{\alpha})^3\Gamma_w(\lfloor\frac{n}{\alpha}\rfloor). \tag{22}$$

Note that

$$\left\lfloor \frac{\lfloor \frac{n}{\beta_1}\rfloor}{\beta_2}\right\rfloor \le \left\lfloor \frac{n}{\beta_1\beta_2}\right\rfloor,$$

where $\beta_1, \beta_2 \ge 1$ are real constants. Then the inequality (22) implies that

$$\Gamma_w(\lfloor\frac{n}{\alpha^j}\rfloor) \le 2q^2(\frac{n}{\alpha^{j+1}})^3\Gamma_w(\lfloor\frac{n}{\alpha^{j+1}}\rfloor). \tag{23}$$

From (22) and (23):

$$\begin{aligned}\Gamma_w(n) \le 2q^2(\frac{n}{\alpha})^3\Gamma_w(\lfloor\frac{n}{\alpha}\rfloor) \le 2q^2(\frac{n}{\alpha})^3 2q^2(\frac{n}{\alpha^2})^3\Gamma_w(\lfloor\frac{n}{\alpha^2}\rfloor) \le \\ 2q^2(\frac{n}{\alpha})^3 2q^2(\frac{n}{\alpha^2})^3 2q^2(\frac{n}{\alpha^3})^3\Gamma_w(\lfloor\frac{n}{\alpha^3}\rfloor) \le \cdots \le \\ \left(\prod_{j\ge 1}^{\lfloor\frac{\ln n}{\ln\alpha}\rfloor} 2q^2(\frac{n}{\alpha^j})^3\right)\Gamma_w(2).\end{aligned} \tag{24}$$

Realize that

$$\frac{n}{\alpha^{\lfloor\frac{\ln n}{\ln\alpha}\rfloor}} \ge 1 \text{ and } \frac{n}{\alpha^{\lceil\frac{\ln n}{\ln\alpha}\rceil}} \le 1.$$

Knowing that $\Gamma_w(2) = q$ and using Lemma 2.25 we obtain from (24):

$$\Gamma_w(n) \le (2q^2)^{\frac{\ln n}{\ln\alpha}}\left(n^{\frac{\ln n}{2\ln\alpha}}\right)^3\Gamma_w(2) \le q(2q^2n)^{\frac{3\ln n}{2\ln\alpha}}.$$

This ends the proof. ☐

From Lemma 2.19 and Corollary 2.26 it follows easily:

Corollary 2.27. *If $w \in R$ and $n > 0$ then*

$$|\operatorname{Pal}_w(n)| \le nq(2q^2n)^{\delta\ln n}.$$

Remark 2.28. Although Corollary 2.26 requires $n > 2$, it is easy to verify that Corollary 2.27 holds also for $n = \{1, 2\}$. That is why we define $n > 0$ in Corollary 2.27.

We can simply apply the upper bound for the palindromic complexity to construct an upper bound for the factor complexity:

Corollary 2.29. *If $w \in R$ and $n > 0$ then*

$$|\operatorname{Fac}_w(n)| \le n^4q^2(2q^2n)^{2\delta\ln n}.$$

Proof. We apply again the property of rich words that every factor is determined by its longest palindromic prefix and its longest palindromic suffix [7]. If there are at most t palindromic factors in w of length $\leq n$, then clearly there can be at most t^2 different factors of length n. Let $\hat{\mathrm{Pal}}_w(k) = \max\{|\mathrm{Pal}_w(j)| \mid 0 \leq j \leq k\}$. From Corollary 2.27 we can deduce that

$$t \leq \sum_{i=1}^{n} |\mathrm{Pal}_w(i)| \leq n\hat{\mathrm{Pal}}_w(n) \leq n^2 q (2q^2 n)^{\log_2 n}.$$

The corollary follows. □

3. Rich words closed under reversal

We can improve our upper bound for the factor complexity if we use the inequality (1). This inequality was shown for infinite words whose set of factors is closed under reversal. The next lemma and proposition generalize the existing proof for finite words $w \in A^+$ with $\mathrm{Fac}_w(n+1)$ closed under reversal.

First we introduce an alphabet B and an infinite word $\kappa(w)$. Let $B = A \cup \{x, y\}$ be an alphabet such that $x, y \notin A$; it follows that $|B| = |A| + 2$. Given $w \in A^+$, let $\kappa(w) = (wxw^R y)^\infty \in B^\infty$.

We show that $\kappa(w)$ preserves richness.

Lemma 3.1. *If $w \in A^+$ is rich then $\kappa(w) \in B^\infty$ is also rich.*

Proof. We have that wx is rich, because w is rich and $\mathrm{lps}(wx) = x$, which is a unioccurrent palindrome in wx and wxw^R is a palindromic closure of the rich word wx, which preserves richness [9]. As well $wxw^R y$ is rich, because y is a unioccurrent palindrome in $wxw^R y$. Suppose that $(wxw^R y)^j$ is rich, where j is a positive integer. We prove that $(wxw^R y)^{j+1}$ is rich.

We have that $\mathrm{lps}(wxw^R y)^j = y(wxw^R y)^{j-1}$ and thus $(wxw^R y)^j wxw^R$ is a palindromic closure which is rich. Realize that $\mathrm{lps}(wxw^R y)^{j+1} = y(wxw^R y)^j$ and $y(wxw^R y)^j$ is unioccurrent in $y(wxw^R y)^{j+1}$. Thus $y(wxw^R y)^{j+1}$ is rich. It follows that all prefixes of $\kappa(w)$ are rich. Since all factors of rich words are rich, we proved that all factors of $\kappa(w)$ are rich. Consequently $\kappa(w)$ is rich. This completes the proof. □

The following proposition generalizes the inequality (1) for finite words. It is known that for rich infinite words whose set of factors is closed under reversal, the inequality may be replaced with equality; this result has been proved in [6]. We prove also the equality for finite rich words.

Proposition 3.2. *If $w \in A^+$, $\mathrm{Fac}_w(n+1)$ is closed under reversal, $|w| \geq n+1$, and $n > 0$ then*

$$|\mathrm{Pal}_w(n)| + |\mathrm{Pal}_w(n+1)| \leq |\mathrm{Fac}_w(n+1)| - |\mathrm{Fac}_w(n)| + 2.$$

If w is also rich then the inequality becomes equality, formally:

$$|\operatorname{Pal}_w(n)| + |\operatorname{Pal}_w(n+1)| = |\operatorname{Fac}_w(n+1)| - |\operatorname{Fac}_w(n)| + 2.$$

Proof. Let $t = \kappa(w)$ and let $k \in \{n, n+1\}$. Clearly if $\operatorname{Fac}_w(n+1)$ is closed under reversal and $i \leq n$ that $\operatorname{Fac}_w(i)$ is also closed under reversal. Thus we have that

$$\begin{aligned}\operatorname{Fac}_t(k) = \operatorname{Fac}_w(k) \cup \{uxv \mid u, v \in A^* \text{ and } u \text{ is a suffix of } w \text{ and}\\ v \text{ is a prefix of } w^R \text{ and } |uxv| = k\} \cup\\ \{uyv \mid u, v \in A^* \text{ and } u \text{ is a suffix of } w^R \text{ and}\\ v \text{ is a prefix of } w \text{ and } |uyv| = k\}.\end{aligned} \tag{25}$$

The formula (25) says that the set of factors of t having length k contains:

- the set of factors of w of length k,
- the set of factors of t containing one occurrence of x, and
- the set of factors of t containing one occurrence of y.

It is easy to see that there are no other factors in $\operatorname{Fac}_t(k)$. Moreover for every $i \in \{0, 1, 2, \ldots, k-1\}$ there are unique $u \in \operatorname{Fac}_w(i)$ and $v \in \operatorname{Fac}_w(k-i-1)$ such that $uxv \in \operatorname{Fac}_t(k)$ ($uyv \in \operatorname{Fac}_t(k)$). It follows that

$$|\operatorname{Fac}_t(k)| = |\operatorname{Fac}_w(k)| + 2k. \tag{26}$$

Obviously t contains exactly two palindromes r_1, r_2 such that r_1, r_2 are not factors of w and $|r_1| = |r_2| \in \{n, n+1\}$. In addition $r_1 = uxu^R$ and $r_2 = vyv^R$ for some words u, v. Formally

$$\begin{aligned}\operatorname{Pal}_t(n+1) \cup \operatorname{Pal}_t(n) = \operatorname{Pal}_w(n+1) \cup \operatorname{Pal}_w(n) \cup \{uxu^R, vyv^R \mid u \text{ is a suffix of } w \text{ and}\\ v \text{ is a suffix of } w^R \text{ and } |uxu| = |vyv^R| \in \{n, n+1\}\}\end{aligned}$$

It follows that

$$|\operatorname{Pal}_t(n+1)| + |\operatorname{Pal}_t(n)| = |\operatorname{Pal}_w(n+1)| + |\operatorname{Pal}_w(n)| + 2. \tag{27}$$

Clearly Fac_t is closed under reversal; realize that t has infinitely many palindromic prefixes. Consequently (1) holds for t. Then from (1), (26), and (27) we have that

$$|\operatorname{Pal}_t(n)| + |\operatorname{Pal}_t(n+1)| \leq |\operatorname{Fac}_t(n+1)| - |\operatorname{Fac}_t(n)| + 2$$

and

$$|\operatorname{Pal}_w(n)| + |\operatorname{Pal}_w(n+1)| + 2 \leq |\operatorname{Fac}_w(n+1)| + 2(n+1) - |\operatorname{Fac}_w(n)| - 2n + 2. \tag{28}$$

It follows from (28) that

$$|\operatorname{Pal}_w(n)| + |\operatorname{Pal}_w(n+1)| \leq |\operatorname{Fac}_w(n+1)| - |\operatorname{Fac}_w(n)| + 2.$$

If w is rich then Lemma 3.1 implies that t is rich. Then it follows from [6], (26), and (27) that

$$|\operatorname{Pal}_t(n)| + |\operatorname{Pal}_t(n+1)| = |\operatorname{Fac}_t(n+1)| - |\operatorname{Fac}_t(n)| + 2$$

and

$$|\operatorname{Pal}_w(n)| + |\operatorname{Pal}_w(n+1)| + 2 = |\operatorname{Fac}_w(n+1)| + 2(n+1) - |\operatorname{Fac}_w(n)| - 2n + 2. \quad (29)$$

It follows from (29) that

$$|\operatorname{Pal}_w(n)| + |\operatorname{Pal}_w(n+1)| = |\operatorname{Fac}_w(n+1)| - |\operatorname{Fac}_w(n)| + 2.$$

This completes the proof. □

Based on Proposition 3.2 we can present a new relation for palindromic and factor complexity.

Proposition 3.3. *Let* $\hat{\operatorname{Pal}}_w(k) = \max\{|\operatorname{Pal}_w(j)| \mid 0 \leq j \leq k\}$. *If* $w \in R$ *is a rich word such that* $\operatorname{Fac}_w(n+1)$ *is closed under reversal,* $|w| \geq n+1$, *and* $n > 0$, *then*

$$|\operatorname{Fac}_w(n)| \leq 2(n-1)\hat{\operatorname{Pal}}_w(n) - 2(n-1) + q.$$

Proof. Proposition 3.2 states for rich words that

$$|\operatorname{Pal}_w(n)| + |\operatorname{Pal}_w(n+1)| - 2 = |\operatorname{Fac}_w(n+1)| - |\operatorname{Fac}_w(n)|. \quad (30)$$

Since $\operatorname{Fac}_w(n+1)$ closed under reversal, we have that $\operatorname{Fac}_w(i)$ is closed under reversal for $i \leq n+1$. We can sum (30) over all lengths $i \leq n$:

$$\sum_{i=1}^{n-1}(|\operatorname{Pal}_w(i)| + |\operatorname{Pal}_w(i+1)| - 2) = \sum_{i=1}^{n-1}(|\operatorname{Fac}_w(i+1)| - |\operatorname{Fac}_w(i)|). \quad (31)$$

The sums from (31) may be expressed as follows:

$$\begin{aligned}\sum_{i=1}^{n-1}(|\operatorname{Fac}_w(i+1)| - |\operatorname{Fac}_w(i)|) &= \operatorname{Fac}_w(2) - \operatorname{Fac}_w(1) + \operatorname{Fac}_w(3) - \operatorname{Fac}_w(2) \\ &\quad + \operatorname{Fac}_w(4) - \operatorname{Fac}_w(3) + \cdots + \operatorname{Fac}_w(n-1) - \operatorname{Fac}_w(n-2) \\ &\quad + \operatorname{Fac}_w(n) - \operatorname{Fac}_w(n-1) = \operatorname{Fac}_w(n) - \operatorname{Fac}_w(1).\end{aligned} \quad (32)$$

$$\sum_{i=1}^{n-1}(|\operatorname{Pal}_w(i)| + |\operatorname{Pal}_w(i+1)| - 2) \leq (n-1)(\hat{\operatorname{Pal}}_w(n-1) + \hat{\operatorname{Pal}}_w(n) - 2). \quad (33)$$

From (31), (32), and (33) we get:

$$\operatorname{Fac}_w(n) - \operatorname{Fac}_w(1) \leq (n-1)(\hat{\operatorname{Pal}}_w(n-1) + \hat{\operatorname{Pal}}_w(n) - 2).$$

It follows that

$$\mathrm{Fac}_w(n) \le (n-1)(2\hat{\mathrm{Pal}}_w(n) - 2) + \mathrm{Fac}_w(1).$$

This can be reformulated as:

$$\mathrm{Fac}_w(n) \le 2(n-1)\hat{\mathrm{Pal}}_w(n) - 2(n-1) + \mathrm{Fac}_w(1).$$

Since $\mathrm{Fac}_w(1) = q$ it follows that

$$\mathrm{Fac}_w(n) \le 2(n-1)\hat{\mathrm{Pal}}_w(n) - 2(n-1) + q.$$

This completes the proof. □

Proposition 3.3 and Lemma 2.27 imply an improvement to our upper bound for the factor complexity for rich words with $\mathrm{Fac}_w(n+1)$ closed under reversal:

Corollary 3.4. *If $w \in R$ with $\mathrm{Fac}_w(n+1)$ closed under reversal, $|w| \ge n+1$, and $n > 0$, then:*

$$|\,\mathrm{Fac}_w(n)| \le 2(n-1)nq(2q^2n)^{\delta \ln n} - 2(n-1) + q.$$

Since the palindromic closure of finite rich words is closed under reversal, we can improve the upper bound for factor complexity for finite rich words.

Corollary 3.5. *If $w \in R$ and $n > 0$ then*

$$|\,\mathrm{Fac}_w(n)| \le 2(2n-1)2nq(4q^2n)^{\delta \ln 2n} - 2(2n-1) + q.$$

Proof. Palindromic closure $\hat{w}$ of a word $w \in R$ preserves richness. Furthermore $\mathrm{Fac}_{\hat{w}}$ is closed under reversal, $\mathrm{Fac}_w \subseteq \mathrm{Fac}_{\hat{w}}$, and $|\tilde{w}| \le 2|w|$ [9]. Hence we can apply Corollary 3.4, where we replace n with $2n$. □

Theorem 1.1 in the introduction presents a "simple" (although a somewhat worse) upper bound for the factor complexity. Here follows the proof.

Proof of Theorem 1.1. Note that for $n > 0$ we have that

$$2(2n-1)2nq(4q^2n)^{\delta \ln 2n} - 2(2n-1) + q \le 8n^2q(4q^2n)^{\delta \ln 2n} \le (4q^2n)^{\delta \ln 2n + 2}.$$

The theorem follows from Corollary 3.5. □

Acknowledgments

The author wishes to thank to Štěpán Starosta for his useful comments. The author acknowledges support by the Czech Science Foundation grant GAČR 13-03538S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/205/OHK4/3T/14.

References

[1] J.-P. Allouche, M. Baake, J. Cassaigne, and D. Damanik, *Palindrome complexity*, Theor. Comput. Sci., 292 (2003), pp. 9–31.

[2] P. Balá̌zi, Z. Masáková, and E. Pelantová, *Factor versus palindromic complexity of uniformly recurrent infinite words*, Theor. Comput. Sci., 380 (2007), pp. 266–275.

[3] L. Balková, *Beta-integers and Quasicrystals*, PhD thesis, Czech Technical University in Prague and Université Paris Diderot-Paris 7, 2008.

[4] L. Balková, E. Pelantová, and Š. Starosta, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO-Theor. Inf. Appl., 44 (2010), pp. 443–470.

[5] A. Blondin Massé, S. Brlek, S. Labbé, and L. Vuillon, *Palindromic complexity of codings of rotations*, Theor. Comput. Sci., 412 (2011), pp. 6455–6463.

[6] M. Bucci, A. De Luca, A. Glen, and L. Q. Zamboni, *A connection between palindromic and factor complexity using return words*, Adv. in Appl. Math., 42 (2009), pp. 60–74.

[7] ———, *A new characteristic property of rich words*, Theor. Comput. Sci., 410 (2009), pp. 2860–2863.

[8] 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.

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

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

[11] D. Kosolobov, M. Rubinchik, and A. M. Shur, *Pal^k is linear recognizable online*, in SOFSEM 2015: Theory and Practice of Computer Science, G. F. Italiano, T. Margaria-Steffen, J. Pokorný, J.-J. Quisquater, and R. Wattenhofer, eds., Berlin, Heidelberg, 2015, Springer Berlin Heidelberg, pp. 289–301.

[12] E. Pelantová and Š. Starosta, *On words with the zero palindromic defect*, in Combinatorics on Words, S. Brlek, F. Dolce, C. Reutenauer, and É. Vandomme, eds., Cham, 2017, Springer International Publishing, pp. 59–71.

[13] E. A. Petrova and A. M. Shur, *Transition property for cube-free words*, in Computer Science – Theory and Applications, R. van Bevern and G. Kucherov, eds., Cham, 2019, Springer International Publishing, pp. 311–324.

[14] M. Rubinchik and A. M. Shur, *The number of distinct subpalindromes in random words*, Fund. Inform., 145 (2016), pp. 371–384.

[15] 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.

[16] ———, *Transition property for α-power free languages with $\alpha \geq 2$ and $k \geq 3$ letters*, in Developments in Language Theory, N. Jonoska and D. Savchuk, eds., Cham, 2020, Springer International Publishing, pp. 294–303.

[17] L. Schaeffer and J. Shallit, *Closed, palindromic, rich, privileged, trapezoidal, and balanced words in automatic sequences*, Electr. J. Comb., 23 (2016), p. P1.25.

[18] J. Shallit and A. Shur, *Subword complexity and power avoidance*, Theoretical Computer Science, 792 (2019), pp. 96 – 116. Special issue in honor of the 70th birthday of Prof. Wojciech Rytter.

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

Communicated by (The editor will be set by the publisher).
September 29, 2020.

Article [[Ru04]]: Rich Words Containing Two Given Factors

Rich Words Containing Two Given Factors

Josef Rukavicka(✉)

Department of Mathematics,
Faculty of Nuclear Sciences and Physical Engineering,
Czech Technical University in Prague,
Trojanova 13, 120 01 Prague 2, Czech Republic
josef.rukavicka@seznam.cz

Abstract. A finite word w with $|w| = n$ contains at most $n+1$ distinct palindromic factors. If the bound $n+1$ is attained, the word w is called *rich*. Let $\mathrm{F}(w)$ be the set of factors of the word w. It is known that there are pairs of rich words that cannot be factors of a same rich word. However it is an open question how to decide for a given pair of rich words u, v if there is a rich word w such that $\{u, v\} \subseteq \mathrm{F}(w)$. We present a response to this open question:

If w_1, w_2, w are rich words, $m = \max\{|w_1|, |w_2|\}$, and $\{w_1, w_2\} \subseteq \mathrm{F}(w)$ then there exists also a rich word $\bar{w}$ such that $\{w_1, w_2\} \subseteq \mathrm{F}(\bar{w})$ and $|\bar{w}| \leq m2^{k(m)+2}$, where $k(m) = (q+1)m^2(4q^{10}m)^{\log_2 m}$ and q is the size of the alphabet. Hence it is enough to check all rich words of length equal or lower to $m2^{k(m)+2}$ in order to decide if there is a rich word containing factors w_1, w_2.

1 Introduction

In the last years there have appeared several articles dealing with rich words; see, for instance, [1–3,5]. Recall that a palindrome is a word that reads the same forwards and backwards, for example "noon" and "level". If a word w of length n contains $n+1$ distinct palindromic factors then the word w is called *rich*. It is known that a word of length n can contain at most $n+1$ palindromic factors including the empty word. The notion of a rich word has been extended also to infinite words. An infinite word is called rich if its every finite factor is rich [3,4].

Let $\mathrm{lps}(w)$ and $\mathrm{lpp}(w)$ denote the longest palindromic suffix and the longest palindromic prefix of a word w, respectively. The authors of [1] showed the following property of rich words:

Proposition 1. *If r, t are two factors of a rich word w such that $\mathrm{lps}(r) = \mathrm{lps}(t)$ and $\mathrm{lpp}(r) = \mathrm{lpp}(t)$, then $r = t$.*

Two related open questions can be found:

– In [5]: Is the condition in Proposition 1 sufficient for two rich words u and v to be factors of the same rich word?

R. Mercaş and D. Reidenbach (Eds.): WORDS 2019, LNCS 11682, pp. 286–298, 2019.
https://doi.org/10.1007/978-3-030-28796-2_23

– In [3]: We do not know how to decide whether two rich words u and v are factors of a same rich word w.

In the current article we present a response to the question from [3] in the following form: We prove that if w_1, w_2, w are rich words, $m = \max\{|w_1|, |w_2|\}$, and $\{w_1, w_2\} \subseteq \mathrm{F}(w)$ then there exists a rich word $\bar{w}$ such that $\{w_1, w_2\} \subseteq \mathrm{F}(\bar{w})$ and $|\bar{w}| \leq m2^{k(m)+2}$, where $k(m) = (q+1)m^2(4q^{10}m)^{\log_2 m}$ and q is the size of the alphabet. Thus it is enough to check all rich words of length equal or lower to $m2^{k(m)+2}$ in order to decide if there is a rich word containing factors w_1, w_2. However it is a rather theoretic way how to check the existence of such a word, since the number of words needed to be checked grows "pretty rapidly" with the length of the factors in question.

We describe the basic ideas of the proof. If w is a rich word, then let a be a letter such that $\mathrm{lps}(wa) = a\,\mathrm{lpps}(w)a$, where lpps denotes the longest proper palindromic suffix. It is known and easy to show that wa is a rich word [5, Proof of Theorem 2.1]. Thus every rich word w can be richly extended to a word wa. We will call wa a *standard extension* of w. If there is a letter b such that $a \neq b$ and wb is also a rich word, then we call the longest palindromic suffix of wb a *flexed palindrome*; the explication of the terminology is that wb is not a standard extension of w, hence wb is "flexed" from the standard extension. We define a set Γ of pairs of rich words (w, r), where r is a flexed palindrome of w, the longest palindromic prefix of w does not contain the factor r, and $|r| \geq |\bar{r}|$ for each flexed palindrome $\bar{r}$ of w. If $(w, r) \in \Gamma$, w_1 is the prefix of w with $|w_1| = |r| - 1$ and w_2 is the suffix of w with $|w_2| = |r| - 1$ then we construct a rich word $\bar{w}$ possessing the following properties:

– The word w_1 is a prefix of $\bar{w}$ and the word w_2 is a suffix of $\bar{w}$.
– The number of occurrences of r in $\bar{w}$ is strictly smaller than the number of occurrences of r in w.
– The set of flexed palindromes of $\bar{w}$ is a subset of the set of flexed palindromes of w.

Iterative applying of this construction will allow us for a given rich word w with a prefix w_1 and a suffix w_2 to construct a rich word t containing factors w_1, w_2 and having no flexed palindrome longer than m, where $m = \max\{|w_1|, |w_2|\}$.

Another important, but simple, observation is that if w is a rich word with prefix u such that the number of flexed palindromes in w is less than k and u has exactly one occurrence in w then there is an upper bound for the length of w. We show this upper bound as a function of k and consequently we derive an upper bound for the length of t.

2 Preliminaries

Let A be a finite alphabet with $q = |\,\mathrm{A}\,|$. The elements of A will be called letters. Let ϵ denote the empty word.
Let A^* be the set of all finite words over A including the empty word and let $\mathrm{A}^n \subset \mathrm{A}^*$ be the set of all words of length n.

Let $\mathrm{R} \subset \mathrm{A}^*$ denote the set of all rich words.
Let $\mathrm{F}(w) \subset \mathrm{A}^*$ denote the set of all factors of $w \in \mathrm{A}^*$; we state explicitly that $\epsilon, w \in \mathrm{F}(w)$. Let $\mathrm{F}(S) = \bigcup_{v \in S} \mathrm{F}(v)$, where $S \subseteq \mathrm{A}^*$.
Let $\mathrm{F}_p(w) \subseteq \mathrm{F}(w)$ be set of all palindromic factors of $w \in \mathrm{A}^*$.
Let $\mathrm{Prf}(w)$ and $\mathrm{Suf}(w)$ be the set of all prefixes and all suffixes of $w \in A^*$ respectively; we define that $\{\epsilon, w\} \subseteq \mathrm{Prf}(w) \cap \mathrm{Suf}(w)$.
Let w^R denote the reversal of $w \in A^*$; formally if $w = w_1 w_2 \dots w_k$ then $w^R = w_k \dots w_2 w_1$, where $w_i \in \mathrm{A}$ and $i \in \{1, 2, \dots, k\}$. In addition we define that $\epsilon^R = \epsilon$.

Let $\mathrm{lps}(w)$ and $\mathrm{lpp}(w)$ denote the longest palindromic suffix and the longest palindromic prefix of $w \in \mathrm{A}^*$ respectively. We define that $\mathrm{lps}(\epsilon) = \mathrm{lpp}(\epsilon) = \epsilon$.
Let $\mathrm{lpps}(w)$ and $\mathrm{lppp}(w)$ denote the longest proper palindromic suffix and the longest proper palindromic prefix of $w \in \mathrm{A}^*$ respectively, where $|w| \geq 2$.
Let $\mathrm{trim}(w) = v$, where $v, w \in \mathrm{A}^*$, $x, y \in \mathrm{A}$, $w = xvy$, and $|w| \geq 2$.
Let $\mathrm{rtrim}(w) = v$, where $v, w \in \mathrm{A}^*$, $y \in \mathrm{A}$, $w = vy$, and $|w| \geq 1$.
Let $\mathrm{ltrim}(w) = v$, where $v, w \in \mathrm{A}^*$, $x \in \mathrm{A}$, $w = xv$, and $|w| \geq 1$.

Example 2. If $\mathrm{A} = \{1, 2, 3, 4, 5\}$ and $w = 124135$, then $\mathrm{trim}(w) = 2413$, $\mathrm{ltrim}(w) = 24135$, and $\mathrm{rtrim}(w) = 12413$.

Let $\mathrm{pc}(w)$ be the palindromic closure of $w \in \mathrm{A}^*$; formally $\mathrm{pc}(w) = uvu^R$, where $w = uv$ and $v = \mathrm{lps}(w)$. Note that $\mathrm{pc}(w)$ is a palindrome.

Let $\mathrm{MinLenWord}(U)$ and $\mathrm{MaxLenWord}(U)$ be the shortest and the longest word from the set U respectively, where either $U \subseteq \mathrm{Prf}(w)$ or $U \subseteq \mathrm{Suf}(w)$ for some $w \in A^*$. If $U = \emptyset$ then we define $\mathrm{MinLenWord}(U) = \epsilon$ and $\mathrm{MaxLenWord}(U) = \epsilon$.

Let $\mathrm{lcp}(w_1, w_2)$ be the longest common prefix of words $w_1, w_2 \in \mathrm{A}^*$; formally $\mathrm{lcp}(w_1, w_2) = \mathrm{MaxLenWord}(\mathrm{Prf}(w_1) \cap \mathrm{Prf}(w_2))$.
Let $\mathrm{lcs}(w_1, w_2)$ be the longest common suffix of words $w_1, w_2 \in \mathrm{A}^*$; formally $\mathrm{lcs}(w_1, w_2) = \mathrm{MaxLenWord}(\mathrm{Suf}(w_1) \cap \mathrm{Suf}(w_2))$.
Let $\mathrm{occur}(u, v)$ be the number of occurrences of v in u, where $u, v \in \mathrm{A}^*$ and $|v| > 0$; formally $\mathrm{occur}(u, v) = |\{w \mid w \in \mathrm{Suf}(u) \text{ and } v \in \mathrm{Prf}(w)\}|$. We call a factor v *unioccurrent* in u if $\mathrm{occur}(u, v) = 1$.

Recall the notion of a *complete return* [2]: Given a word w and factors $r, u \in \mathrm{F}(w)$, we call the factor r a complete return to u in w if r contains exactly two occurrences of u, one as a prefix and one as a suffix.

We list some known properties of rich words that we use in our article. All of them can be found, for instance, in [2].

Proposition 3. *If $w, u \in \mathrm{R}$, $|w| \geq 1$, $|u| \geq 1$, and $u \in \mathrm{F}_p(w)$ then all complete returns to u in w are palindromes.*

Proposition 4. *If $w \in \mathrm{R}$ and $p \in \mathrm{F}(w)$ then $p, p^R \in \mathrm{R}$.*

Proposition 5. *A word w is rich if and only if every prefix $p \in \mathrm{Prf}(w)$ has a unioccurrent palindromic suffix.*

3 Standard Extensions and Flexed Palindromes

We start with a formal definition of a standard extension and a flexed palindrome introduced at the beginning of the article.

Definition 6. *Let $j \geq 0$ be a nonnegative integer, $w \in \mathrm{R}$, and $|w| \geq 2$. We define* $\mathrm{StdExt}(w, j)$ *as follows:*

- $\mathrm{StdExt}(w, 0) = w$.
- $\mathrm{StdExt}(w, 1) = wa$ *such that* $\mathrm{lps}(wa) = a\,\mathrm{lpps}(w)a$ *and* $a \in \mathrm{A}$.
- $\mathrm{StdExt}(w, j) = \mathrm{StdExt}(\mathrm{StdExt}(w, j-1), 1)$, *where* $j > 1$.

Let $\mathrm{StdExt}(w) = \{\mathrm{StdExt}(w, j) \mid j \geq 0\}$. *If* $p \in \mathrm{StdExt}(w)$ *then we call* p *a standard extension of* w.
Let $\mathrm{T}(w) = \{\mathrm{lps}(ub) \mid ub \in \mathrm{Prf}(w)$ *and* $b \in \mathrm{A}$ *and* $ub \neq \mathrm{StdExt}(u, 1)\}$. *If* $r \in \mathrm{T}(w)$ *then we call* r *a flexed palindrome of* w.

For a given rich word $w \in \mathrm{R}$ having a flexed palindrome r we define a *standard palindromic replacement* of r to be the longest palindromic suffix of a standard extension of a prefix p of w such that $\mathrm{lps}(px) = r$, where px is a prefix of w and $x \in \mathrm{A}$. The idea is that we can "replace" r with the standard palindromic replacement.

Definition 7. *Let* $\mathrm{stdPalRep}(w, r) = \mathrm{lps}(\mathrm{StdExt}(p, 1))$, *where* $w, r \in \mathrm{R}$, $r \in \mathrm{T}(w)$, $px \in \mathrm{Prf}(w)$, $x \in \mathrm{A}$, *and* $\mathrm{lps}(px) = r$.
We call $\mathrm{stdPalRep}(w, r)$ *a standard palindromic replacement of* r *in* w.

Example 8. If $\mathrm{A} = \{0, 1\}$ and $w = 1101011\mathbf{001100}11$ then $001100 \in \mathrm{T}(w)$, $\mathrm{lps}(1101011\mathbf{001100}) = 001100$, $\mathrm{StdExt}(110101100110, 1) = 1101011001101$, and $\mathrm{stdPalRep}(w, 001100) = \mathrm{lps}(1101011001101) = 1011001101$.

We show that the length of a flexed palindrome r is less than the length of the standard palindromic replacement $\mathrm{stdPalRep}(w, r)$.

Lemma 9. *If* $ux, uy \in \mathrm{R}$, $x, y \in \mathrm{A}$, $x \neq y$, *and* $ux = \mathrm{StdExt}(u, 1)$ *then* $|\mathrm{lps}(ux)| > |\mathrm{lps}(uy)|$.

Proof. Let $yty = \mathrm{lps}(uy)$. From the definition of a standard extension we have $\mathrm{lps}(ux) = xvx$, where $v = \mathrm{lpps}(u)$ and hence $t \in \mathrm{Suf}(v)$. Since $y \neq x$ we have also $yt \in \mathrm{Suf}(v)$. The lemma follows.

An obvious corollary is that a flexed palindrome of w is not a prefix of w.

Corollary 10. *If* $w, r \in \mathrm{R}$ *and* $r \in \mathrm{T}(w)$ *then* $r \notin \mathrm{Prf}(w)$.

In [5] the standard extension has been used to prove that each rich word w can be extended "richly"; this means that there is $a \in A$ such that wa is rich.

Lemma 11. *If* $w \in \mathrm{R}$ *and* $|w| \geq 2$ *then* $\mathrm{StdExt}(w) \subset R$.

Proof. Obviously it is enough to prove that $\mathrm{StdExt}(w, 1) \in \mathrm{R}$, since for every $t \in \mathrm{StdExt}(w) \setminus \{w\}$ there is a rich word $\bar{t}$ such that $t = \mathrm{StdExt}(\bar{t}, 1)$.

Let $xpx = \mathrm{lps}(\mathrm{StdExt}(w, 1))$, where $x \in \mathrm{A}$. Proposition 5 implies that we need to prove that xpx is unioccurrent in $\mathrm{StdExt}(w, 1)$. Realize that p is unioccurrent in w, hence xpx is unioccurrent in $\mathrm{StdExt}(w, 1)$.

To simplify the proofs of the paper we introduce a function $\mathrm{MaxStdExt}(u, v)$ to be the longest prefix z of u such that z is also a standard extension of v:

Definition 12. *Let* $\mathrm{MaxStdExt}(u, v) = \mathrm{MaxLenWord}(\{\mathrm{StdExt}(v) \cap \mathrm{Prf}(u)\})$, *where* $u \in \mathrm{R}$ *and* $v \in \mathrm{Prf}(u)$. *We call* $\mathrm{MaxStdExt}(u, v)$ *a* maximal standard extension *of* v *in* u.

The next lemma shows that if a rich word contains factors ypx and ypy, where p is a palindrome, p is not a prefix of w, x, y are distinct letters, and ypx "occurs" before ypy in w then ypy is a flexed palindrome.

Lemma 13. *If* $w, v, p \in \mathrm{R}$, $v \in \mathrm{Prf}(w)$, $p \notin \mathrm{Prf}(w)$, $x, y \in \mathrm{A}$, $x \neq y$, $ypx \in \mathrm{Suf}(v)$, $ypy \notin \mathrm{F}(v)$, *and* $ypy \in \mathrm{F}(w)$ *then* $ypy \in \mathrm{T}(w)$.

Proof. Let $\bar{v}$ be such that $\bar{v}y \in \mathrm{Prf}(w)$, $ypy \in \mathrm{Suf}(\bar{v}y)$, and $\mathrm{occur}(\bar{v}y, ypy) = 1$. Let $u = \mathrm{lps}(\bar{v})$. Because $p \notin \mathrm{Prf}(w)$ it follows that $u = \mathrm{lpps}(\bar{v}) = \mathrm{lps}(\bar{v})$ and thus there is $z \in \mathrm{A}$ such that $zu \in \mathrm{Suf}(\bar{v})$. Obviously $v \in \mathrm{Prf}(\bar{v})$ and hence $\mathrm{occur}(\bar{v}, p) > 1$. Then Proposition 5 implies that $\mathrm{occur}(u, p) > 1$. It follows that $yp \in \mathrm{Suf}(u) \cap \mathrm{Prf}(u)$, $z \neq y$, and Lemma 9 implies that $ypy \in \mathrm{T}(w)$. The word w with is its factors is depicted on Fig. 1. This completes the proof.

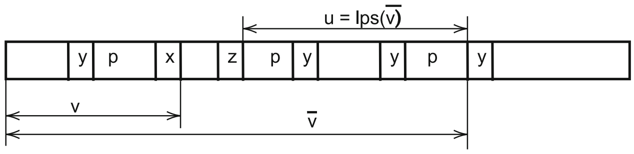

Fig. 1. Structure of the word w for Lemma 13.

4 Removing Flexed Points

We define formally the set Γ mentioned in the introduction. An element (w, r) of the set Γ represents a rich word w for which we are able to construct a new rich word $\bar{w}$ such that $\bar{w}$ does not contain the flexed palindrome r, but $\bar{w}$ have certain common prefixes and suffixes with w. We require that r is one of the longest flexed palindromes of w and that r is not a factor of the longest palindromic prefix of w. In addition we require that $|r| > 2$ so that the standard extension of $\mathrm{rtrim}(r)$ would be defined.

Definition 14. *Let Γ be a set defined as follows:* $(w, r) \in \Gamma$ *if*

1. $w, r \in \mathrm{R}$ *and* $|r| > 2$ *and* $r \in \mathrm{T}(w)$ *and*
2. $r \notin \mathrm{F}(\mathrm{lpp}(w))$ *and*
3. $|r| \geq |\bar{r}|$ *for each* $\bar{r} \in \mathrm{T}(w)$.

Given $(w, r) \in \Gamma$, we need to express w as a concatenation of its factors having some special properties. For this reason we define a function $\mathrm{parse}(w, r)$:

Definition 15. *If* $(w, r) \in \Gamma$ *then let* $\mathrm{parse}(w, r) = (v, z, t)$, *where*

- $v, z, t \in \mathrm{R}$ *and* $vzt = w$ *and*
- $r \in \mathrm{Suf}(v)$ *and* $\mathrm{occur}(w, r) = \mathrm{occur}(v, r)$ *and*
- $vz = \mathrm{MaxStdExt}(vzt, v)$.

Remark 16. The prefix v is the shortest prefix of w that contains all occurrences of r. The prefix vz is the maximal standard extension of v in w, and t is such that $vzt = w$. It is easy to see that v, z, t exist and are uniquely determined for $(w, r) \in \Gamma$.

The next simple lemma is necessary for the following definition of a *reduced prefix*.

Lemma 17. *Let* $(w, r) \in \Gamma$, *let* $(v, z, t) = \mathrm{parse}(w, r)$, *and let* $\bar{v}$ *be such that* $v = \bar{v}\,\mathrm{lps}(v)$.

- *If* $\mathrm{occur}(\bar{v}r, r) > 1$ *then there is a word* $\bar{g}$ *such that* $\bar{g}rz \in \mathrm{Prf}(v)$ *and* $\mathrm{occur}(\bar{g}rz, r) < \mathrm{occur}(v, r)$
- *If* $\mathrm{occur}(\bar{v}r, r) = 1$ *then* $U \neq \emptyset$ *and* $r \notin \mathrm{F}(U)$, *where* $U = \{u \mid u \in \mathrm{Prf}(\mathrm{pc}(\bar{v}\,\mathrm{rtrim}(r)))$ *and* $\mathrm{ltrim}(r)z \in \mathrm{Suf}(u)\}$.

Proof. It follows from Property 2 of Definition 14 that there is $h \in \mathrm{Prf}(w)$ such that $w = hz^R\,\mathrm{lps}(v)zt$. Note that $\mathrm{lps}(v) \neq v$ since $r \in \mathrm{T}(w)$ and thus $r \notin \mathrm{Prf}(w)$, see Corollary 10. It is clear that $r \in \mathrm{Prf}(\mathrm{lps}(v)) \cap \mathrm{Suf}(\mathrm{lps}(v))$. This implies that $hz^Rr \in \mathrm{Prf}(w)$. Note that $\bar{v} = hz^R$. We distinguish two cases as stated in the Lemma:

- $\mathrm{occur}(\bar{v}r, r) > 1$: Let g be the complete return to r in v such that $g \in \mathrm{Suf}(hz^Rr)$. Clearly $rz \in \mathrm{Prf}(g)$ and $z^Rr \in \mathrm{Suf}(g)$, since $r \notin \mathrm{F}(\mathrm{ltrim}(r)z)$; recall $r \in \mathrm{Suf}(v)$ and $\mathrm{occur}(v, r) = \mathrm{occur}(vzt, r)$. Let $\bar{g}$ be such that $\bar{g}g = hz^Rr$.
- If $\mathrm{occur}(\bar{v}r, r) = 1$: Let $\bar{u} = \mathrm{stdPalRep}(hz^Rr, r)$. Clearly $\mathrm{lps}(hz^Rr) = r$ and $\bar{u} \neq r$. Because $z^R\,\mathrm{rtrim}(r) \in \mathrm{Suf}(hz^R\,\mathrm{rtrim}(r))$, then obviously $U \neq \emptyset$ and $r \notin \mathrm{F}(U)$.

The word w with is its factors is depicted on Fig. 2. This completes the proof.

For an element $(w, r) \in \Gamma$ we define a function $\operatorname{rdcPrf}(w, r)$ (the reduced prefix), which is a prefix of the palindromic closure of some prefix of w. In Theorem 28 we show that the concatenation of $\operatorname{rdcPrf}(w, r)$ and t is a rich word having a strictly smaller number of occurrences of r than in w, where $(v, z, t) = \operatorname{parse}(w, r)$. This reducing of occurrences of r is the key for removing all "long" flexed palindromes as explained in the introduction.

Definition 18. *If* $w, r \in \Gamma$ *and* $(v, z, t) = \operatorname{parse}(w, r)$ *then let* $\operatorname{rdcPrf}(w, r)$ *be defined as follows. Following the notation and the proof of Lemma 17 we distinguish two cases:*

- $\operatorname{occur}(\bar{v}r, r) > 1$*: We define* $\operatorname{rdcPrf}(w, r) = \bar{g}rz$.
- $\operatorname{occur}(\bar{v}r, r) = 1$*: We define* $\operatorname{rdcPrf}(w, r) = \operatorname{MinLenWord}(U)$.

We call $\operatorname{rdcPrf}(w, r)$ *the reduced prefix of* w *by* r.

Figure 2 depicts the factors of the word w used for construction of the reduced prefix of w.

Remark 19. Note in Definition 18 in the second case, where $\operatorname{occur}(\bar{v}r, r) = 1$, it may happen that the reduced prefix $\operatorname{rdcPrf}(w, r)$ is not a prefix of w. However it is a prefix of a palindromic closure of $hz^R \operatorname{rtrim}(r)$, hence the number of flexed palindromes remains the same; formally $|\operatorname{T}(hz^R \operatorname{rtrim}(r))| = |\operatorname{T}(\operatorname{rdcPrf}(w, r))|$. Realize that $\operatorname{pc}(t) \in \operatorname{StdExt}(t)$ for each $t \in \mathrm{R}$ and $|t| \geq 2$.

In the first case, where $\operatorname{occur}(\bar{v}r, r) > 1$, the reduced prefix $\operatorname{rdcPrf}(w, r)$ is always a prefix of w.

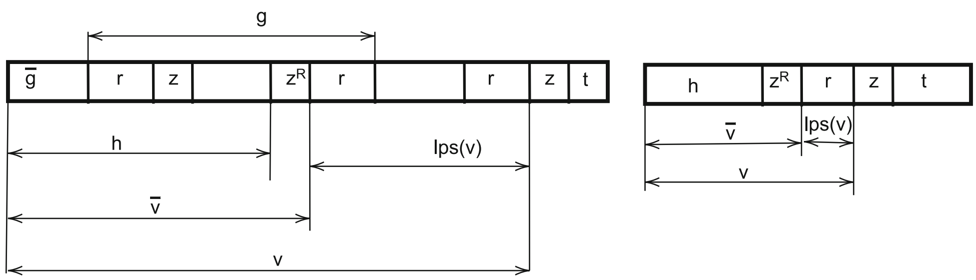

Fig. 2. Construction of the reduced prefix. Case 1 and 2.

To clarify the definition of the reduced prefix $\operatorname{rdcPrf}(w, r)$ we present below two examples representing those two cases in the definition. For both examples we consider that $\mathrm{A} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Example 20. If $w = 1239993223999324423999322255223993$ and $r = 999$ then $v = 1239993223999324423999$, $z = 322$, $t = 55223993$, $\operatorname{lps}(v) = 999324423999$, $h = 1239993$, $w = hz^R \operatorname{lps}(v) zt$, $g = 9993223999 \in \operatorname{Suf}(hz^R r) = \operatorname{Suf}(1239993223999)$, $\bar{g} = 123$, and $\operatorname{rdcPrf}(w, r) = 123999322$.

Example 21. If $w = 123999599932239949$ and $r = 999$ then $v = 1239995999$, $z = 32$, $t = 239949$, $\mathrm{lps}(v) = 9995999$, $h = 1$, $w = hz^R \mathrm{lps}(v)zt$, $\mathrm{StdExt}(hz^R \mathrm{rtrim}(r), 1) = \mathrm{StdExt}(12399, 1) = 123993$, $\bar{u} = \mathrm{stdPalRep}(123999, 999) = 3993$, $\mathrm{pc}(12399) = 12399321$, $U = \{1239932\}$, and $\mathrm{rdcPrf}(w, r) = 1239932$.

Using the reduced prefix we can now define the word $\mathrm{rdcWrd}(w, r)$ (a *reduced word*):

Definition 22. *Let* $\mathrm{rdcWrd}(w, r) = \mathrm{rdcPrf}(w, r)t$*, where* $(v, z, t) = \mathrm{parse}(w, r)$ *and* $(w, r) \in \Gamma$*. We call* $\mathrm{rdcWrd}(w, r)$ *the reduced word of* w *by* r*.*

We show that the reduced word $\mathrm{rdcWrd}(w, r)$ and w have the same prefix and suffix of length $|r| - 1$.

Lemma 23. *If* $(w, r) \in \Gamma$ *and* $u = \mathrm{rdcWrd}(w, r)$ *then* $|\mathrm{lcp}(u, w)| \geq |r| - 1$ *and* $|\mathrm{lcs}(u, w)| \geq |r| - 1$*.*

Proof. From the construction of the reduce prefix and the reduced word, it is easy to see that $\mathrm{rtrim}(r) \in \mathrm{F}(\mathrm{lcp}(u, v))$ and $\mathrm{ltrim}(r) \in \mathrm{F}(\mathrm{lcs}(u, v))$. The lemma follows.

As already mentioned the reduced prefix $\mathrm{rdcPrf}(w, r)$ is not necessarily a prefix of w. In such a case $\mathrm{rdcPrf}(w, r) \in \mathrm{Prf}(\mathrm{pc}(\bar{v}\, \mathrm{rtrim}(r)))$, see Definition 18. We show that every palindrome from the set $\mathrm{F}(\mathrm{rdcPrf}(w, r)) \setminus \mathrm{F}(\bar{v}\, \mathrm{rtrim}(r)))$ contains as a factor the standard palindromic replacement $\bar{u}$ of r in w and we show that $\bar{u}$ is not a factor of w. This will be important when proving richness of the word $\mathrm{rdcWrd}(w, r)$.

Let $\mathrm{F}(w, r) = \{u \mid u \in \mathrm{F}(w) \text{ and } r \notin \mathrm{F}(u)\} \subseteq \mathrm{F}(w)$, where $w, r \in \mathrm{A}^*$. The set $\mathrm{F}(w, r)$ contains factors of w that do not contain the factor r. Let $\mathrm{F}_p(w, r) = \mathrm{F}_p(w) \cap \mathrm{F}(w, r)$.

Proposition 24. *If* $(w, r) \in \Gamma$*,* $(v, z, t) = \mathrm{parse}(w, r)$*,* $u = \mathrm{rdcPrf}(w, r)$*,* $\bar{u} = \mathrm{stdPalRep}(w, r)$*, and* $\bar{v}$ *is such that* $v = \bar{v}\, \mathrm{lps}(v)$ *then* $\mathrm{F}_p(u, \bar{u}) \subseteq \mathrm{F}_p(\bar{v}\, \mathrm{rtrim}(r))$ *and* $\bar{u} \notin \mathrm{F}_p(w)$*.*

Proof. From the properties of the palindromic closure it is easy to see that $\mathrm{F}_p(\mathrm{pc}(f), \mathrm{lps}(f)) \subseteq \mathrm{F}_p(f)$ for each $f \in \mathrm{R}$. It means that every palindromic factor of $\mathrm{pc}(f)$ that is not a factor of f contains the factor $\mathrm{lps}(f)$. It follows that $\mathrm{F}_p(u, \bar{u}) \subseteq \mathrm{F}_p(\mathrm{rtrim}(v))$.

We show that $\mathrm{occur}(w, \bar{u}) = 0$. Let $\bar{u} = xtx$ and $r = ypy$, where $x, y \in \mathrm{A}$. Obviously $x \neq y$. Lemma 9 implies that $|\bar{u}| > |r|$. It follows that $py \in \mathrm{Prf}(t)$, and $yp \in \mathrm{Suf}(t)$. Thus $xty \in \mathrm{F}(w)$. Lemma 13 implies that $\bar{u} \in \mathrm{F}_p(w)$ if and only if $\bar{u} \in \mathrm{T}(w)$. Since $|\bar{u}| > |r|$, this would be a contradiction to Property 3 of Definition 14. Hence $\bar{u} \notin \mathrm{F}_p(w)$. This completes the proof.

We define a set Mergeable which contains 3-tuples (d, g, t) of rich words such that, among other properties, dg and gt are rich. Later we prove that the "merge" dgt of dg and gt is also rich. Let $\mathrm{flt}(p) = \mathrm{A} \cap \mathrm{Prf}(p)$ be the first letter of a word $p \in A^*$ with $|p| \geq 1$.

Definition 25. *We define a set* Mergeable *as follows:* $(d, g, t) \in \text{Mergeable}$ *if*

1. $d, g, t, dg, gt, dg\,\text{flt}(t) \in \text{R}$ *and*
2. $\text{lps}(dg\,\text{flt}(t)) \in \text{T}(dg\,\text{flt}(t))$ *and*
3. $\text{lps}(gp) \notin \text{F}(dg)$ *for each* $p \in \text{Prf}(t)$ *with* $|p| \geq 1$.

Let $(d, g, t) \in \text{Mergeable}$. The following proposition shows that dgt is a rich word. This will allow us from a rich word of the form $dgwgt$ to construct a rich word dgt. In other words this will allow us to remove the factor w from a rich word, and thus to reduce the number of occurrences of flexed palindromes.

Proposition 26. *If* $(d, g, t) \in \text{Mergeable}$ *then*

- $dgt \in \text{R}$ *and*
- $\text{lps}(dgp) = \text{lps}(gp)$ *for each* $p \in \text{Prf}(t)$ *with* $|p| \geq 1$.

Proof. From Definition 25 it follows immediately that the Proposition holds for $(d, g, \text{flt}(t))$.

Suppose that the Proposition holds for $(d, g, \bar{p})$, where $\bar{p} \in \text{Prf}(t)$ with $1 \leq |\bar{p}| < |t|$. We show that the Proposition holds for (d, g, p) and (h, g, p), where $p \in \text{Prf}(t)$ with $|p| = |\bar{p}|+1$. From the property that a finite rich word w of length n has $n+1$ palindromic factors it follows that $|\,\text{F}_p(w)| = |\,\text{F}_p(\text{rtrim}(w))|+1$. This and Property 3 of Definition 25 imply that $\text{lps}(gp) \notin \text{F}(\text{lps}(dg\bar{p}))$. Consequently $\text{lps}(gp) = \text{lps}(dgp)$ and $dgp \in \text{R}$, see Proposition 5. This completes the proof.

We prove that the set of flexed palindromes of the word dgt that are not factors of prefix dg, where $(d, g, t) \in \text{Mergeable}$, does not depend on the prefix d.

Proposition 27. *If* $(d, g, t), (h, g, t) \in \text{Mergeable}$*,* $|d| \geq 1$*, and* $|h| \geq 1$ *then* $\text{T}(dgt) \setminus \text{T}(dg) = \text{T}(hgt) \setminus \text{T}(hg)$.

Proof. To get a contradiction, suppose that there is $p \in \text{Prf}(t)$ with $|p| \geq 1$ such that $\text{lps}(dgp) \in \text{T}(dgp)$ and $\text{lps}(hgp) \notin \text{T}(hgp)$. If $|p| > 1$ then $|\,\text{lps}(dgp)| \leq |\,\text{lps}(dg\,\text{rtrim}(p))|$ and $\text{trim}(\text{lps}(hgp)) = \text{lps}(hg\,\text{rtrim}(p))$, which is a contradiction, because $\text{lps}(dg\,\text{rtrim}(p)) = \text{lps}(hg\,\text{rtrim}(p)) = \text{lps}(g\,\text{rtrim}(p))$, see Proposition 26. If $|p| = 1$ the proposition holds because of Property 2 of Definition 25. This completes the proof.

The main theorem of the paper states that the reduced word $\text{rdcWrd}(w, r)$ is rich, where $(w, r) \in \Gamma$. In addition the theorem asserts that the set of flexed palindromes of $\text{rdcWrd}(w, r)$ is a subset of the set of flexed palindromes of the word w, the number of occurrences of r is strictly smaller in $\text{rdcWrd}(w, r)$ than in w, and the longest common prefix and suffix of $\text{rdcWrd}(w, r)$ and w are longer than $|r| - 1$.

Theorem 28. *If* $(w, r) \in \Gamma$ *then*

- $\text{rdcWrd}(w, r) \in \text{R}$ *and* $\text{T}(\text{rdcWrd}(w, r)) \subseteq \text{T}(w)$ *and*
- $\text{occur}(\text{rdcWrd}(w, r), r) < \text{occur}(w, r)$ *and*
- $|\,\text{lcp}(\text{rdcWrd}(w, r), w)| \geq |r| - 1$ *and* $|\,\text{lcs}(\text{rdcWrd}(w, r), w)| \geq |r| - 1$.

Proof. Recall that $\mathrm{rdcWrd}(w,r) = ut$, where $(v,z,t) = \mathrm{parse}(w,r)$ and $u = \mathrm{rdcPrf}(w,r)$. If $|t| = 0$ then $\mathrm{rdcWrd}(w,r) \in \mathrm{R}$ and $\mathrm{T}(\mathrm{rdcWrd}(w,r)) \subseteq \mathrm{T}(w)$.

Let d be such that $\mathrm{rdcPrf}(w,r) = d\,\mathrm{ltrim}(r)z$. If $|t| > 0$ then we are going to show that $(d, \mathrm{ltrim}(r)z, t) \in \mathrm{Mergeable}$. Obviously $d\,\mathrm{ltrim}(r)z, \mathrm{ltrim}(r)zt \in \mathrm{R}$; recall that $\mathrm{ltrim}(r)zt \in \mathrm{Suf}(w)$. We need to show that Property 3 of Definition 25 is satisfied: Because $vz = \mathrm{MaxStdExt}(vzt, v)$ it follows that $\mathrm{lps}(vz\,\mathrm{flt}(t)) \in \mathrm{T}(w)$. This and $\mathrm{occur}(\mathrm{ltrim}(r)zt, r) = 0$ imply that $|\,\mathrm{lps}(vzp)| \leq |\,\mathrm{ltrim}(r)zp|$ for each $p \in \mathrm{Prf}(t)$ with $|p| \geq 1$. In consequence $\mathrm{lps}(\mathrm{ltrim}(r)zp) = \mathrm{lps}(vzp)$. Proposition 24 and $\mathrm{occur}(vzp, \mathrm{lps}(vzp)) = 1$ imply that $\mathrm{lps}(vzp) \notin \mathrm{F}(d\,\mathrm{ltrim}(r)z)$. The other properties of Definition 25 are clearly also fulfilled. Hence $(d, \mathrm{ltrim}(r)z, t) \in \mathrm{Mergeable}$. Thus from Proposition 26 we get that $d\,\mathrm{ltrim}(r)zt \in \mathrm{R}$.

Let $\bar{w}$ be such that $w = \bar{w}\,\mathrm{ltrim}(r)zt$. Obviously $(\bar{w}, \mathrm{ltrim}(r)z, t) \in \mathrm{Mergeable}$. Then Proposition 27 asserts that $\mathrm{T}(\mathrm{rdcWrd}(w,r)) \subseteq \mathrm{T}(w)$.

The fact that $\mathrm{occur}(ut, r) < \mathrm{occur}(w, r)$ follows Lemma 17 and Definition 18. Note that $\mathrm{occur}(\mathrm{rdcPrf}(w,r), r) < \mathrm{occur}(w,r)$.

The properties $|\,\mathrm{lcp}(\mathrm{rdcWrd}(w,r), w)| \geq |r| - 1$ and $|\,\mathrm{lcs}(\mathrm{rdcWrd}(w,r), w)| \geq |r| - 1$ follow from Lemma 23.

This completes the proof.

Two more examples will illuminate the construction of $\mathrm{rdcWrd}(w,r)$. The examples are again based on the two cases of Definition 18. For both example we consider that $\mathrm{A} = \{1,2,3,4,5,6,7,8\}$.

Example 29. If $w = 1214565654774565654565656547874$ and $r = 656$ then $v = 121456565477456565456 56$, $z = 547$, $t = 874$, $\mathrm{lps}(v) = 656545656$, $u = \mathrm{rdcPrf}(w,r) = 12145656547$, and $\mathrm{rdcWrd}(w,r) = ut = 12145656547874$.

Example 30. If $w = 12145656547874$ and $r = 656$ then $v = 12145656$, $z = 54$, $t = 7874$, $\mathrm{lps}(v) = 656$, $u = \mathrm{rdcPrf}(w,r) = 12145654$, and $\mathrm{rdcWrd}(w,r) = ut = 121456547874$.

For a finite set S, we can consider that the set S is well-ordered. No matter how, we just need a function that selects one element from S. Let the function $\mathrm{selectFirst}(S)$ returns the first element of S. If S is an empty set, then we define $\mathrm{selectFirst}(S) = \epsilon$.

If a rich word w has a factor u, then the palindromic closure of w is rich and contains the factor u^R. Hence for us when constructing a rich word containing given factors, it does not matter if w contains u or u^R. We introduce the notion of a *reverse-unioccurrent* factor. Moreover we define a function $\mathrm{ruo}(w,u,v)$ (a *reverse-unioccurrence* of u,v in w) which returns a factor of w such that u,v are reverse-unioccurrent in this factor; in addition we require u or u^R to be a prefix and v or v^R to be a suffix of $\mathrm{ruo}(w,u,v)$.

Definition 31. *If* $|\{u,u^R\} \cap \mathrm{F}(w)| = 1$ *then we say that a word* u *is reverse-unioccurrent in* w*, where* $w,u \in \mathrm{R}$.

If $w_1, w_2, w \in \mathrm{R}$*,* $w_1, w_2 \in \mathrm{F}(w)$*, and there is* $t \in \mathrm{Prf}(w)$ *such that* $w_1 \in \mathrm{F}(t)$ *and* $\{w_2, w_2^R\} \cap \mathrm{F}(t) = \emptyset$ *then let* $\mathrm{M}(w,w_1,w_2) \subset \mathrm{F}(w)$ *such that* $t \in \mathrm{M}(w,w_1,w_2)$ *if:*

– *$t \in \mathrm{F}(w)$ and w_1, w_2 are reverse-unioccurrent in t and*
– *$\{w_1, w_1^R\} \cap \mathrm{Prf}(t) \neq \emptyset$ and $\{w_2, w_2^R\} \cap \mathrm{Suf}(t) \neq \emptyset$.*

Let $\mathrm{ruo}(w, w_1, w_2) = \mathrm{selectFirst}(\mathrm{M}(w, w_1, w_2))$*.*

Remark 32. It is not difficult to see that the function $\mathrm{ruo}(r, w_1, w_2)$ is well defined and the set $\mathrm{M}(w, w_1, w_2)$ is nonempty.

We define the function $\mathrm{elmWrd}(w, w_1, w_2)$ (*eliminated word*) that constructs a rich word from w by "eliminating all" flexed palindromes longer than $m = \max\{|w_1|, |w_2|\}$ and keeping the prefix w_1 and the suffix w_2 of w.

Definition 33. *Let* $\mathrm{maxFlxPal}(w) = \{r \mid (w, r) \in \Gamma\}$*. If $w, w_1, w_2 \in \mathrm{R}$, $m = \max\{|w_1|, |w_2|\}$, $w_1 \in \mathrm{Prf}(w)$, and $w_2 \in \mathrm{Suf}(w)$, then let* $\mathrm{elmWrd}(w, w_1, w_2)$ *be the result of the following procedure:*

```
01 INPUT: w,m,w_1,w_2;
02 res: = ruo(w,w_1,w_2);
03 r := selectFirst(maxFlxPal(res));
04 WHILE r is longer than m
05 DO
06  res := rdcWrd(res,r);
07  res := ruo(res,w_1,w_2);
08  r := selectFirst(maxFlxPal(res));
09 END-DO;
10 RETURN res;
```

The calls of the function ruo on the lines 02 and 07 guarantee that w_1, w_2 are reverse-unioccurrent in the word res and that $\{w_1, w_1^R\} \cap \mathrm{Prf}(res) \neq \emptyset$ and $\{w_2, w_2^R\} \cap \mathrm{Suf}(res) \neq \emptyset$. Realize that it is not guaranteed that w_1, w_2 are reverse-unioccurrent in $\mathrm{rdcWrd}(res, r)$, even if w_1, w_2 are reverse-unioccurrent in res.

Clearly, the facts that $\bar{t}$ is reverse-unioccurrent in a rich word t and $\bar{t} \in \mathrm{Prf}(t)$ imply that $\mathrm{lpp}(t) \in \mathrm{Prf}(\bar{t})$; realize that if $d \in \mathrm{F}(\mathrm{lpp}(\bar{t}))$ then $d^R \in \mathrm{F}(\mathrm{lpp}(t))$ also, since palindromes are closed under reversal. Thus if r is a flexed palindrome of t longer than the prefix $\bar{t}$, then r is not a factor of $\mathrm{lpp}(t)$ and hence r satisfies Property 2 of Definition 14.

Let $r = \mathrm{selectFirst}(\mathrm{maxFlxPal}(w))$. The call of the function $\mathrm{rdcWrd}(res, r)$ on the line 06 contains valid parameters, since if $r \neq \epsilon$ and $|r| > m$ then $(w, r) \in \Gamma$.

In addition, because $|r| > \max\{|w_1|, |w_2|\}$, Theorem 28 asserts that $\{w_1, w_1^R\} \cap \mathrm{Prf}(\mathrm{rdcWrd}(res, r)) \neq \emptyset$ and $\{w_2, w_2^R\} \cap \mathrm{Suf}(\mathrm{rdcWrd}(res, r)) \neq \emptyset$; consequently $\{w_1, w_1^R\} \cap \mathrm{Prf}(res) \neq \emptyset$ and $\{w_2, w_2^R\} \cap \mathrm{Suf}(res) \neq \emptyset$ on the line 06.

Moreover Theorem 28 implies that the procedure finishes after a finite number of iterations, because $\mathrm{occur}(\mathrm{rdcWrd}(w, r), r) < \mathrm{occur}(w, r)$ and $\mathrm{T}(\mathrm{rdcWrd}(w, r)) \subseteq \mathrm{T}(w)$. The number of iterations is bounded by the number $\sum_{r \in \mathrm{T}(w)} \mathrm{occur}(w, r)$. Note that several occurrences of r may be "eliminated" in one iteration. Hence we proved the following lemma:

Lemma 34. *If* $w \in \mathrm{R}$*,* $w_1 \in \mathrm{Prf}(w)$*,* $w_2 \in \mathrm{Suf}(w)$*,* $m = \max\{|w_1|, |w_2|\}$*, and* $t = \mathrm{elmWrd}(w, w_1, w_2)$ *then*

– $t \in \mathrm{R}$ *and for each* $r \in \mathrm{T}(t)$ *we have* $|r| \leq m$ *and*
– $\{w_1, w_1^R\} \cap \mathrm{Prf}(t) \neq \emptyset$ *and* $\{w_2, w_2^R\} \cap \mathrm{Suf}(t) \neq \emptyset$*.*

5 Words with Limited Number of Flexed Points

What is the maximal length of a word u such that w is reverse-unioccurrent in u, w is a prefix of u, and u has a given maximal number of flexed palindromes? The proposition below answers this question.

Proposition 35. *If* $u, w \in \mathrm{R}$*,* $|u| \geq 1$*,* $|v| \geq 1$*,* $w \in \mathrm{Prf}(u)$*,* $|\,\mathrm{T}(u) \setminus \mathrm{T}(w)| \leq k$*,* $|w| \leq m$*, and* w *is reverse-unioccurrent in* u *then* $|u| \leq m2^{k+1}$*.*

Proof. Obviously $|\,\mathrm{pc}(u)| < 2|u|$, $\mathrm{pc}(u) \in \mathrm{StdExt}(u)$, and w is not reverse-unioccurrent in $\mathrm{pc}(u)$, since $w^R \in \mathrm{Suf}(\mathrm{pc}(u))$. It follows that if $v_1, v_2 \in \mathrm{Prf}(\bar{u})$ such that v_1 is reverse-unioccurrent in $\bar{u}$, $v_1 \in \mathrm{Prf}(v_2)$, $|\,\mathrm{T}(v_2) \setminus \mathrm{T}(v_1)| = 1$, and $\mathrm{lps}(v_2) \in \mathrm{T}(v_2)$ then $|\,\mathrm{rtrim}(v_2)| < 2|v_1|$, since $\mathrm{rtrim}(v_2) \in \mathrm{StdExt}(v_1)$ and $\mathrm{pc}(v_1) \in \mathrm{StdExt}(v_1)$ also. This implies that $|v_2| \leq 2|v_1|$. The proposition follows.

Remark 36. The proof asserts that if v_1, v_2 are two prefixes of a word u such that the longest palindromic suffix of v_2 is the only flexed palindrome in v_2 which is not a factor of v_1, then v_2 is at most twice longer than v_1 on condition that v_1 is reverse-unioccurrent in $\mathrm{ltrim}(v_2)$. Less formally it means that the length of a word can grow at most twice before the next flexed palindrome appears. Note that for $k = 1$ we have $|u| \leq 2m$, which makes sense, since the palindromic closure of a word u is at most twice longer than u.

In [4] the author showed an upper bound for the number of palindromic factors of given length in a rich word. Recall that $q = |\,\mathrm{A}\,|$.

Proposition 37 ([4], Corollary 2.23]). *If* $w \in \mathrm{R}$ *and* $n > 0$ *then*

$$|F_p(w) \cap \mathrm{A}^n\,| \leq (q+1)n(4q^{10}n)^{\log_2 n}.$$

Proposition 37 implies an upper bound for the number of flexed palindromes:

Lemma 38. *If* $w \in \mathrm{R}$*,* $n > 0$*, and* $\mathrm{A}^{\leq n} = \bigcup_{j=0}^{n} \mathrm{A}^j$ *then*

$$|\,\mathrm{T}(w) \cap \mathrm{A}^{\leq n}\,| \leq (q+1)n^2(4q^{10}n)^{\log_2 n}.$$

Proof. Just realize that $\sum_{j=1}^{n}(q+1)j(4q^{10}j)^{\log_2 j} \leq (q+1)n^2(4q^{10}n)^{\log_2 n}$.

From Lemmas 34, 38 and Proposition 35 we obtain the result of the article:

Corollary 39. *If* w, w_1, w_2 *are rich words,* $w_1, w_2 \in \mathrm{F}(w)$*,* $m = \max\{|w_1|, |w_2|\}$ *then there exists also a rich word* $\bar{w}$ *such that* $w_1, w_2 \in \mathrm{F}(\bar{w})$ *and* $|\bar{w}| \leq m2^{k(m)+2}$*, where* $k(m) = (q+1)m^2(4q^{10}m)^{\log_2 m}$*.*

Proof. Without loss of generality, suppose that there is $\bar{t} \in \mathrm{Prf}(w)$ such that $w_1 \in \mathrm{Prf}(\bar{t})$ and $\{w_2, w_2^R\} \cap \mathrm{F}(\bar{t}) = \emptyset$. Then the function $\mathrm{ruo}(w, w_1, w_2)$ is well-defined. Let $t \in \mathrm{ruo}(w, w_1, w_2)$. Consider the word $g = \mathrm{elmWrd}(t, w_1, w_2)$. Let $k(m) = (q+1)m^2(4q^{10}m)^{\log_2 m}$. Lemma 38 and Proposition 35 imply that $|g| \geq m2^{k(m)+1}$. Lemma 34 implies that $g \in \mathrm{R}$, $\{w_1, w_1^R\} \cap \mathrm{F}(g) \neq \emptyset$, and $\{w_2, w_2^R\} \cap \mathrm{F}(g) \neq \emptyset$. Let $\bar{w} = \mathrm{pc}(g)$. It follows that $w_1, w_2 \in \mathrm{F}(\bar{w})$. Because $|\,\mathrm{pc}(g)| \leq 2|g|$, the corollary follows.

Acknowledgments. The author wishes to thank to Štěpán Starosta for his useful comments. The author acknowledges support by the Czech Science Foundation grant GAČR 13-03538S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/205/OHK4/3T/14.

References

1. Bucci, M., De Luca, A., Glen, A., Zamboni, L.Q.: A new characteristic property of rich words. Theor. Comput. Sci. **410**, 2860–2863 (2009)
2. Glen, A., Justin, J., Widmer, S., Zamboni, L.Q.: Palindromic richness. Eur. J. Comb. **30**, 510–531 (2009)
3. Pelantová, E., Starosta, Š.: On words with the zero palindromic defect. In: Brlek, S., Dolce, F., Reutenauer, C., Vandomme, É. (eds.) WORDS 2017. LNCS, vol. 10432, pp. 59–71. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66396-8_7
4. Rukavicka, J.: An upper bound for palindromic and factor complexity of rich words. (2018, submitted for publication). https://arxiv.org/abs/1810.03573
5. Vesti, J.: Extensions of rich words. Theor. Comput. Sci. **548**, 14–24 (2014)

Article [[Ru05]]: A Unique Extension of Rich Words

A Unique Extension of Rich Words

Josef Rukavicka*

October 01, 2019
Mathematics Subject Classification: 68R15

Abstract

A word w is called rich if it contains $|w|+1$ palindromic factors, including the empty word. We say that a rich word w can be extended in at least two ways if there are two distinct letters x, y such that wx, wy are rich.

Let R denote the set of all rich words. Given $w \in \mathrm{R}$, let $\mathrm{K}(w)$ denote the set of all words such that if $u \in \mathrm{K}(w)$ then $wu \in \mathrm{R}$ and wu can be extended in at least two ways. Let $\omega(w) = \min\{|u| \mid u \in \mathrm{K}(w)\}$ and let $\phi(n) = \max\{\omega(w) \mid w \in \mathrm{R} \text{ and } |w| = n\}$, where $n > 0$. Vesti (2014) showed that $\phi(n) \leq 2n$. In other words, it says that for each $w \in \mathrm{R}$ there is a word u with $|u| \leq 2|w|$ such that $wu \in \mathrm{R}$ and wu can be extended in at least two ways.

We prove that $\phi(n) \leq n$. In addition we prove that for each real constant $c > 0$ and each integer $m > 0$ there is $n > m$ such that $\phi(n) \geq (\frac{2}{9} - c)n$. The results hold for each finite alphabet having at least two letters.

1 Introduction

A word is called a *palindrome* if it is equal to its reversal. Two examples of palindromes are "noon" and "level". It is known that a word w can contain at most $|w|+1$ distinct palindromic factors, including the empty word [2]. If the bound $|w|+1$ is attained, the word w is called *rich*. Quite many

*Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, CZECH TECHNICAL UNIVERSITY IN PRAGUE (josef.rukavicka@seznam.cz).

articles investigated the properties of rich words in recent years, for example [1, 2, 3, 4, 5]. Some of the properties of rich words are stated in the next section; see Propositions 2.1, 2.2, and 2.3.

In [3] it was proved that if w is rich then there is a letter x such that wx is also rich. In [5] it was proved that if w is rich then there is a word u and two distinct letters x, y such that $|u| \leq 2|w|$ and wux, wuy are rich. Concerning this result, the author of [5] formulated an open question:

- Let w be a rich word. How long is the shortest u such that wu can always be extended in at least two ways?

In the current article we improve the result from [5] and as such, to some extent, we answer to the open question. Let R denote the set of all rich words. We say that a rich word w *can be extended in at least two ways* if there are two distinct letters x, y such that wx, wy are rich. Given $w \in \mathrm{R}$, let $\mathrm{K}(w)$ denote the set of all words such that if $u \in \mathrm{K}(w)$ then $wu \in \mathrm{R}$ and wu can be extended in at least two ways; $\mathrm{K}(w)$ contains the empty word if w can be extended in at least two ways. Let $\omega(w) = \min\{|u| \mid u \in \mathrm{K}(w)\}$ and let $\phi(n) = \max\{\omega(w) \mid w \in \mathrm{R} \text{ and } |w| = n\}$, where $n > 0$. The result from [5] can be presented as $\phi(n) \leq 2n$.

We show that $\phi(n) \leq n$. It is natural to ask how good this bound is. The rich word wu is called a *unique rich extension* of w if there is no proper prefix $\bar{u}$ of u such that $w\bar{u}$ can be extended in at least two ways. In Remark 2.4 in [5] there is an example which shows that there are $w_n, u_n \in \mathrm{R}$ such that $w_n u_n$ is a unique rich extension of w_n and $|u_n| = n$, where $n > 1$. However in the given example the length of w_n grows significantly more rapidly than the length of u_n as n tends towards infinity. This could suggest that $\lim_{n\to\infty} \frac{\phi(n)}{n} = 0$; we show that this suggestion is false. We prove that for each real constant $c > 0$ and each integer $m > 0$ there is $n > m$ such that $\phi(n) \geq (\frac{2}{9} - c)n$.

We explain the idea of the proof. Let w^R denote the reversal of the word w. We construct rich words $\mathrm{h}_n = u_n v^R t v_n$, where $n \geq 3$ such that

1. The word t is the longest palindromic suffix of $u_n v_n^R t$.
2. For every factor xpy of tv_n we have that xpx is a factor of u_n, where x, y are distinct letters and p is a palindrome.
3. $2|\mathrm{h}_n| < |\mathrm{h}_{n+1}|$.

Let $\bar{v}x$ be a prefix of v_n, where x is a letter. Let y be a letter distinct from x and let ypy be the longest palindromic suffix of $u_n v_n^R t\bar{v}y$. Property 1 implies that ypy is a suffix of $y\bar{v}^R t\bar{v}y$, since $\bar{v}^R t\bar{v}$ is the longest palindromic suffix of $u_n v_n^R t\bar{v}$. Property 2 implies that ypy is not unioccurrent in $u_n v_n^R t\bar{v}y$. In consequence $u_n v_n^R t\bar{v}y$ is not rich; see Proposition 2.3. Hence there is no proper prefix v of v_n such that $u_n v_n^R tv$ can be extended in at least two ways. It follows that $|v_n| \leq \omega(u_n v^R t)$. Property 3 implies that for each $m > 0$ there is n such that $|\,\mathrm{h}_n\,| > m$.

We will see that to find u_n for given v_n is quite straightforward. The crucial part of our construction is the word v_n. To be specific, the word v_n that we will present contains only a "small" number of factors xpy defined in Property 2. As a result the length of u_n grows almost linearly with the length of v_n as n tends towards infinity.

2 Preliminaries

Consider an alphabet A with q letters, where $q > 1$. Let A^+ denote the set of all nonempty words over A. Let ϵ denote the empty word, and let $A^* = A^+ \cup \{\epsilon\}$. We have that $\mathrm{R} \subseteq \mathrm{A}^*$.

Let $\mathrm{F}(w)$ be the set of all factors of the word $w \in A^*$; we define that $\epsilon, w \in \mathrm{F}(w)$. Let $\mathrm{Prf}(w)$ and $\mathrm{Suf}(w)$ be the set of all prefixes and all suffixes of $w \in A^*$ respectively; we define that $\{\epsilon, w\} \subseteq \mathrm{Prf}(w) \cap \mathrm{Suf}(w)$.

Let $\mathrm{SufU}(v,u) = \bigcup_{t \in \mathrm{Prf}(u) \setminus \{\epsilon\}} \mathrm{Suf}(vt)$, where $v, u \in \mathrm{A}^*$. The set $\mathrm{SufU}(v,u)$ is the union of sets of suffixes of vt, where t is a nonempty prefix of u.

We define yet the reversal that we have already used in the introduction: Let w^R denote the reversal of $w \in A^*$; formally if $w = w_1 w_2 \dots w_k$ then $w^R = w_k \dots w_2 w_1$, where $w_i \in \mathrm{A}$ and $i \in \{1, 2, \dots, k\}$.

Let $\mathrm{lps}(w)$ and $\mathrm{lpp}(w)$ denote the longest palindromic suffix and the longest palindromic prefix of $w \in \mathrm{A}^*$ respectively. We define that $\mathrm{lps}(\epsilon) = \mathrm{lpp}(\epsilon) = \epsilon$. Let $\mathrm{lpps}(w)$ and $\mathrm{lppp}(w)$ denote the longest proper palindromic suffix and the longest proper palindromic prefix of $w \in \mathrm{A}^*$ respectively, where $|w| \geq 1$. If $|w| = 1$ then we define $\mathrm{lppp}(w) = \mathrm{lpps}(w) = \epsilon$.

Let $\mathrm{rtrim}(w) = v$, where $v, w \in \mathrm{A}^*$, $y \in \mathrm{A}$, $w = vy$, and $|w| \geq 1$. Let $\mathrm{ltrim}(w) = v$, where $v, w \in \mathrm{A}^*$, $x \in \mathrm{A}$, $w = xv$, and $|w| \geq 1$. The functions $\mathrm{rtrim}(w)$ and $\mathrm{ltrim}(w)$ remove the last and the first letter of w respectively.

Let $\mathrm{occur}(u,v)$ be the number of occurrences of v in u, where $u, v \in \mathrm{A}^+$; formally $\mathrm{occur}(u,v) = |\{w \mid w \in \mathrm{Suf}(u) \text{ and } v \in \mathrm{Prf}(w)\}|$. We call a factor

v *unioccurrent* in u if $\operatorname{occur}(u, v) = 1$.

We list some known properties of rich words that we use in our article. All of them can be found, for instance, in [3]. Recall the notion of a *complete return* [3]: Given a word w and factors $r, u \in \mathrm{F}(w)$, we call the factor r a complete return to u in w if r contains exactly two occurrences of u, one as a prefix and one as a suffix.

Proposition 2.1. *If $w, u \in \mathrm{R} \cap \mathrm{A}^+$, $u \in \mathrm{F}(w)$, and u is a palindrome then all complete returns to u in w are palindromes.*

Proposition 2.2. *If $w \in \mathrm{R}$ and $p \in \mathrm{F}(w)$ then $p, p^R \in \mathrm{R}$.*

Proposition 2.3. *A word w is rich if and only if every prefix $p \in \operatorname{Prf}(w)$ has a unioccurrent palindromic suffix.*

From Proposition 2.2 and Proposition 2.3 we have an obvious corollary.

Corollary 2.4. *A word w is rich if and only if every suffix $p \in \operatorname{Suf}(w)$ has a unioccurrent palindromic prefix.*

3 Standard Extension

We define a *left standard extension* and a *right standard extension* of a rich word. The construction of a standard extension has already been used in [5]. The name “standard extension” has been introduced later in [4]. Here we use a different notation and we distinguish a left and a right standard extension.

Definition 3.1. *Let $j \geq 0$ be a nonnegative integer, $w \in \mathrm{R}$, and $|w| \geq 1$. We define $\mathrm{ER}^j(w)$, $\mathrm{EL}^j(w)$ as follows:*

- $\mathrm{ER}^0(w) = \mathrm{EL}^0(w) = w$.
- $\mathrm{EL}(w) = \mathrm{EL}^1(w) = xw$, *where $x \in \mathrm{A}$ is such that* $\operatorname{lppp}(w)x \in \operatorname{Prf}(w)$.
- $\mathrm{ER}(w) = \mathrm{ER}^1(w) = wx$, *where $x \in \mathrm{A}$ is such that* $x \operatorname{lpps}(w) \in \operatorname{Suf}(w)$.
- $\mathrm{EL}^j(w) = \mathrm{EL}(\mathrm{EL}^{j-1}(w))$, *where $j > 1$.*
- $\mathrm{ER}^j(w) = \mathrm{ER}(\mathrm{ER}^{j-1}(w)$, *where $j > 1$.*

Let $\mathrm{EL_a}(w) = \{\mathrm{EL}^j(w) \mid j \geq 0\}$*. We call* $p \in \mathrm{EL_a}(w)$ *a left standard extension of* w*. Let* $\mathrm{ER_a}(w) = \{\mathrm{ER}^j(w) \mid j \geq 0\}$*. We call* $p \in \mathrm{ER_a}(w)$ *a right standard extension of* w*.*

Remark 3.2. It is easy to see that $\mathrm{ER}^j(w) = (\mathrm{EL}^j(w^R))^R$ and $\mathrm{EL}^j(w) = (\mathrm{ER}^j(w^R))^R$, where $j \geq 0$.

If $x \in \mathrm{A}$ then $\mathrm{ER}(x) = \mathrm{EL}(x) = xx$, since $\mathrm{lppp}(x) = \mathrm{lpps}(x) = \epsilon$.

Example 3.3. Let $\mathrm{A} = \{0, 1, 2, 3\}$ and $w = 010200330$. Then we have:

- $\mathrm{lppp}(w) = 010$ and $\mathrm{lpps}(w) = 0330$.
- $\mathrm{ER}(w) = 0102003300$, $\mathrm{ER}^2(w) = 01020033002$, $\mathrm{ER}^3(w) = 010200330020$, $\mathrm{ER}^4(w) = 0102003300201$, $\mathrm{ER}^5(w) = 01020033002010$, $\mathrm{ER}^6(w) = 010200330020102$, $\mathrm{ER}^7(w) = 0102003300201020$.
- $\mathrm{EL}(w) = 2010200330$, $\mathrm{EL}(w)^2 = 02010200330$, $\mathrm{EL}(w)^3 = 002010200330$, $\mathrm{EL}(w)^4 = 3002010200330$, $\mathrm{EL}(w)^5 = 33002010200330$, $\mathrm{EL}(w)^6 = 033002010200330$, $\mathrm{EL}(w)^6 = 0033002010200330$, $\mathrm{EL}(w)^7 = 20033002010200330$.

A left and a right standard extension of a rich word w is rich. In consequence, every rich word w can be extended to rich words wx, yw for some letters x, y; this has already been proved in [3, 4, 5].

Lemma 3.4. *If* $w \in \mathrm{R}$ *and* $|w| \geq 1$ *then* $\mathrm{ER_a}(w) \cup \mathrm{EL_a}(w) \subseteq \mathrm{R}$.

Proof. Since $\mathrm{EL}^j(w) = (\mathrm{ER}^j(w^R))^R$ and since for every $t \in \mathrm{ER_a}(w) \setminus \{w\}$ there is a rich word $\bar{t} \in \mathrm{ER_a}(w)$ such that $t = \mathrm{ER}(\bar{t})$, it is enough to prove that $\mathrm{ER}(w) \in \mathrm{R}$.

Let $xpx = \mathrm{lps}(\mathrm{ER}(w))$, where $x \in \mathrm{A}$. Because $w \in \mathrm{R}$, Proposition 2.3 implies that we need to prove that xpx is unioccurrent in $\mathrm{ER}(w)$. Realize that $p = \mathrm{lpps}(w)$; it means that p is either unioccurrent in w or w is a complete return to p. In either case xpx is unioccurrent in $\mathrm{ER}(w)$. This completes the proof. $\square$

4 A unique rich extension

We formally define a unique rich extension mentioned in the introduction. In addition we define a *flexed point* of a rich word.

Definition 4.1. *If* $u, v \in \mathrm{R} \cap \mathrm{A}^{+}$, $v \in \mathrm{Prf}(u)$, *and*

$$\mathrm{Prf}(\mathrm{rtrim}(u)) \cap \{vt \mid t \in \omega(v)\} = \emptyset$$

then we call u *a* unique rich extension *of* v.

Given $v \in \mathrm{R}$ *with* $|v| > 1$, *let*

$$\mathrm{T}(v) = \{ux \mid ux \in \mathrm{Prf}(v) \text{ and } x \in \mathrm{A} \text{ and } ux \neq \mathrm{ER}(u)\}.$$

We call $w \in \mathrm{T}(v)$ *a flexed point of* v.

Remark 4.2. Note that if $x \in \mathrm{A}$ and ux is a flexed point of a rich word v then u can be extended in at least two ways. A similar notion of a "flexed palindrome" has been used in [4].

Example 4.3. Let $\mathrm{A} = \{0, 1, 2\}$.

- The rich word 00101 can be extended in at least two ways, because 001010, 001011, and 001012 are rich.
- The rich word 20010110 cannot be extended in at least two ways because 200101100 and 200101102 are not rich. Only the right standard extension 200101101 is rich. Hence 200101101 is a unique rich extension of 20010110.
- If $w = 2010110111011110111111001$ then $w1111$ is unique rich extension of w; this example is a modification of the example in Remark 2.4 in [5].
- If $w = 20101101110111110111$ then the set of flexed points of w is:

$$\begin{aligned}\mathrm{T}(w) = \{&20, 201, 20101, 201011, 2010110111,\\ &20101101110111, 201011011101111\}.\end{aligned}$$

There is a connection between a unique rich extension and a right standard extension.

Lemma 4.4. *If* u *is a unique rich extension of* w *then* $u \in \mathrm{ER_a}(w)$.

Proof. Suppose there is $\bar{u}x \in \mathrm{Prf}(u)$ such that $\bar{u} \in \mathrm{ER_a}(w)$, $x \in \mathrm{A}$, and $\bar{u}x \notin \mathrm{ER_a}(w)$. Then obviously $\bar{u}$ can be extended in at least two ways, since both $\bar{u}x$ and $\mathrm{ER}(\bar{u})$ are rich. Hence u cannot be a unique rich extension of w. The lemma follows. □

To simplify the formulation of next lemmas and propositions concerning a unique rich extension we define an auxiliary set Γ as follows: $(v,\bar{v},u)\in\Gamma$ if $v\bar{v}u$ is a unique rich extension of $v\bar{v}$ and $\mathrm{lpps}(v\bar{v})=\bar{v}$, where $v,\bar{v},u\in \mathrm{R}\cap \mathrm{A}^+$.

We show that if wu is unique rich extension of w, then $\mathrm{lpps}(w)$ is unioccurrent in $\mathrm{lpps}(w)u$.

Proposition 4.5. *If* $(v,\bar{v},u)\in\Gamma$ *then* $\mathrm{occur}(\bar{v}u,\bar{v})=1$.

Proof. The proposition follows from the proof of Theorem 2.1 in [5]. The author shows that a rich word w can be extended into a rich word $w\bar{w}$ in such a way that a^n is a suffix of $w\bar{w}$, where a^n is the largest power of some letter $a\in \mathrm{A}$. It is proved that $w\bar{w}$ can be extended in at least two ways. In both cases distinguished in the proof of Theorem 2.1 in [5] it is easy to see that $\mathrm{occur}(\mathrm{lpps}(w)\bar{w},\mathrm{lpps}(w))=1$. The proposition follows. $\square$

We present two simple properties of a unique rich extension.

Lemma 4.6. *Let* $(v,\bar{v},u)\in\Gamma$.

1. *If* $|u|\leq|v|$ *then* $u^R\in\mathrm{Suf}(v)$.
2. *If* $|u|\geq|v|$ *then* $v^R\in\mathrm{Prf}(u)$.

Proof. Obviously $v\bar{v}v^R\in\mathrm{ER_a}(v\bar{v})$. Lemma 4.4 implies that $v\bar{v}u\in\mathrm{ER_a}(v\bar{v})$. The lemma follows. $\square$

The next proposition discusses words of the form $v\bar{v}ux$, where $v\bar{v}ux$ is unique rich extension of $v\bar{v}$, x is a letter, $\bar{v}$ is the longest proper palindromic suffix of $v\bar{v}$, and $\bar{v}ux$ is a flexed point of $\bar{v}ux$. The proposition asserts that there are words t_1,t_2 such that $v=t_1t_2$, xu^R is a proper suffix of t_2, and $\bar{v}t_2^R$ is a flexed point of $\bar{v}t_2^R$. In particular it implies that $|v|>|ux|$.

Proposition 4.7. *If* $(v,\bar{v},ux)\in\Gamma$ *and* $\bar{v}ux\in\mathrm{T}(\bar{v}ux)$ *then there exist* $t_1,t_2\in \mathrm{R}$ *such that*

- $v=t_1t_2$,
- $xu^R\in\mathrm{Suf}(\mathrm{ltrim}(t_2))$, *and*
- $\bar{v}t_2^R\in\mathrm{T}(\bar{v}t_2^R)$.

Proof. Let $w = \mathrm{lpps}(\bar{v}u)$ and let $y \in \mathrm{A}$ be such that $yw \in \mathrm{Suf}(\bar{v}u)$. Since $\bar{v}ux \in \mathrm{T}(\bar{v}ux)$ we have that $x \neq y$.

Obviously $ywy \in \mathrm{F}(v\bar{v}u)$ because $v\bar{v}ux$ is a unique rich extension of $v\bar{v}$ and thus $v\bar{v}uy \notin \mathrm{R}$. Hence the palindromic suffix ywy of $v\bar{v}uy$ is not unioccurrent in $v\bar{v}uy$, see Proposition 2.3.

We have that w is unioccurrent in $\bar{v}u$ and $\bar{v} \notin \mathrm{F}(w)$, since $w = \mathrm{lpps}(\bar{v}u)$ and $\bar{v}$ is unioccurrent in $\bar{v}u$, see Proposition 4.5. It follows that there are $t_1, t_2 \in \mathrm{F}(v)$ such that $v = t_1t_2$, $ywy \in \mathrm{Prf}(t_2\bar{v}ux)$ and ywy is unioccurrent in $t_2\bar{v}ux$. Thus $\mathrm{lpp}(yt_2\bar{v}ux) = ywy$.

From the fact that $\bar{v} \notin \mathrm{F}(w)$ follows that $ywy \in \mathrm{Prf}(t_2\bar{v})$. Lemma 4.6 implies that $|t_2| \geq |ux|$ and $xu^R \in \mathrm{Suf}(\mathrm{ltrim}(t_2))$. Just consider that $|t_2| \leq |ux|$ would imply that $ywy \in \mathrm{F}(\bar{v}ux)$.

Since $xu^R\bar{v} \in \mathrm{Suf}(t_2\bar{v})$, $w \in \mathrm{Suf}(\bar{v}u)$, $ywy \in \mathrm{Prf}(t_2\bar{v})$, and $x \neq y$ it follows that $\mathrm{occur}(t_2\bar{v}, w) > 1$; hence Proposition 2.1 implies that $\mathrm{lppp}(\mathrm{ltrim}(t_2\bar{v})) \neq w$. It follows that $t_2\bar{v} \neq \mathrm{EL}(\mathrm{ltrim}(t_2\bar{v}))$. Consequently $\bar{v}t_2^R \neq \mathrm{ER}(\mathrm{rtrim}(\bar{v}t_2^R))$ and thus $\bar{v}t_2^R \in \mathrm{T}(\bar{v}t_2^R)$. This completes the proof. $\square$

We step to the main result of this section. The theorem says that if $v\bar{v}u$ is a unique rich extension of $v\bar{v}$ and $\bar{v}$ is the longest proper palindromic suffix of $v\bar{v}$ then u is not longer than $v\bar{v}$.

Theorem 4.8. *If* $(v, \bar{v}, u) \in \Gamma$ *then* $|u| \leq |v\bar{v}|$.

Proof. Let $(v, \bar{v}, u) \in \Gamma$. If $|u| + |\mathrm{lpps}(\bar{v})| \leq |\bar{v}|$ then clearly $|u| \leq |v\bar{v}|$. For the rest of the proof suppose that $|u| + |\mathrm{lpps}(\bar{v})| > |\bar{v}|$. We show that the set of flexed points $\mathrm{T}(\bar{v}u)$ is nonempty. Let $\bar{v} = h\,\mathrm{lpps}(\bar{v})$. Proposition 4.5 implies that $h^R \notin \mathrm{Prf}(u)$, because $\mathrm{occur}(h\,\mathrm{lpps}(\bar{v})h^R, \bar{v}) = 2$. Since $|u| + |\mathrm{lpps}(\bar{v})| > |\bar{v}|$ it follows that there are $\bar{u} \in \mathrm{R}$ and $x \in \mathrm{A}$ such that $\bar{u}x \in \mathrm{Prf}(u)$ and $\bar{v}\bar{u}x \neq \mathrm{ER}(\bar{v}\bar{u})$; just realize that $h\,\mathrm{lpps}(\bar{v})h^R \in \mathrm{ER_a}(h\,\mathrm{lpps}(\bar{v}))$. We showed that $\mathrm{T}(\bar{v}u) \setminus \mathrm{Prf}(\bar{v}) \neq \emptyset$.

Without lost of generality, suppose that $\bar{v}\bar{u}x$ is the longest flexed point from the set $\mathrm{T}(\bar{v}u) \setminus \mathrm{Prf}(\bar{v})$ and suppose that $|u| > |v|$. Proposition 4.7 asserts that there are $t_1, t_2 \in \mathrm{R}$ such that $v = t_1t_2$, $\bar{v}t_2^R \in \mathrm{T}(\bar{v}t_2^R)$, and $x\bar{u}^R \in \mathrm{Suf}(\mathrm{ltrim}(t_2))$. If $|u| > |v|$, then $\bar{v}t_2^R \in \mathrm{Prf}(\bar{v}u)$, see Lemma 4.6. This is a contradiction, since we supposed that $\bar{v}\bar{u}x$ is the longest flexed point of $\bar{v}u$. We conclude that $|u| \leq |v|$. This completes the proof. $\square$

The simple corollary is that if wu is a unique rich extension of w then u is not longer than w.

Corollary 4.9. *If $n \geq 1$ then $\phi(n) \leq n$.*

Proof. The corollary is obvious for $n \in \{1, 2\}$. If wu is a unique rich extension of w, $|w| \geq 2$, and $|u| \geq 1$ then there is clearly $(v, \bar{v}, u) \in \Gamma$ such that $w = v\bar{v}$. Then the corollary follows from Theorem 4.8. □

5 Construction of a Uniquely Extensible Rich Word I

Definition 5.1. *We call a word xpy a* switch *if $x, y \in \mathrm{A}$, $x \neq y$, and $p \in \mathrm{A}^*$ is a palindrome. Let $\mathrm{sw}(v) = \{w \mid w \in \mathrm{F}(v)$ and w is a switch$\}$. Let $\mathrm{swSuf}(v, u) = \mathrm{sw}(vu) \cap \mathrm{SufU}(v, u)$, where $v, u \in \mathrm{A}^*$.*

Given $S \subseteq \mathrm{A}^*$, let

$$\mathrm{rdc}(S) = \{w \mid w \in S \text{ and } w \notin \bigcup_{u \in S \setminus \{w\}} \mathrm{F}(u)\}.$$

We call $\mathrm{rdc}(S)$ a reduction *of S.*

Suppose xpy is a switch, let $\mathrm{spc}(xpy) = xpx$, where $x, y \in \mathrm{A}$. We call $\mathrm{spc}(xpy)$ a switch palindromic closure *of the switch xpy. If $B \subset \mathrm{A}^+$ is a set of switches then we define $\mathrm{spc}(B) = \mathrm{rdc}(\bigcup_{w \in B}\{\mathrm{spc}(w)\})$.*

Remark 5.2. Note that if xpy is a switch, then p can be the empty word.

The set $\mathrm{swSuf}(v, u)$ is a set of switches that are suffixes of $v\bar{u}$ for all nonempty prefixes $\bar{u}$ of u.

The reduction $\mathrm{rdc}(S)$ of the set S is a subset of S and contains only elements that are not proper factors of other elements of S.

The switch palindromic closure of a set B is a reduction of the union of all switch palindromic closures of switches from the set B.

Example 5.3. Let $\mathrm{A} = \{0, 1, 2\}$, $v = 0100110$, and $u = 12$. Then we have:

- $\mathrm{sw}(vu) = \{01, 10, 100, 110, 011, 001, 010011, 001101, 12, 012, 11012\}$.
- $\mathrm{swSuf}(v, u) = (\mathrm{sw}(v1) \cap \mathrm{Suf}(v1)) \cup (\mathrm{sw}(v12) \cap \mathrm{Suf}(v12)) =$ $\{001101\} \cup \{12, 012, 11012\}$.
- $\mathrm{spc}(001101) = 001100$, $\mathrm{spc}(12) = 11$, $\mathrm{spc}(012) = 010$, $\mathrm{spc}(11012) = 11011$.

- $\mathrm{spc}(\mathrm{swSuf}(v,u)) = \mathrm{rdc}(\{001100, 11, 010, 11011\}) = \{001100, 010, 110011\}$.

The following proposition clarifies the importance of switches for a unique rich extension of rich words. The proposition says that if

- $wu^R\bar{v}u$ is a rich word and
- $\bar{v}$ is the longest palindromic suffix of $wu^R\bar{v}$ and
- x is a factor of w for every letter and
- for every switch t which is a suffix of $wu^R\bar{v}\bar{u}$ for some $\bar{u} \in \mathrm{Prf}(u)$ we have that $\mathrm{spc}(t)$ is a factor of w

then $wu^R\bar{v}u$ is unique rich extension of $wu^R\bar{v}$.

Proposition 5.4. *If* $w, u, \bar{v} \in \mathrm{A}^+$, $wu^R\bar{v}u \in \mathrm{R}$, $\mathrm{lps}(wu^R\bar{v}) = \bar{v}$, $\mathrm{A} \cap \mathrm{F}(w) = \mathrm{A}$, *and* $\mathrm{spc}(\mathrm{swSuf}(wu^R\bar{v}, u)) \subseteq \mathrm{F}(w)$ *then* $wu^R\bar{v}u$ *is a unique rich extension of* $wu^R\bar{v}$.

Proof. We show that there is no prefix $\bar{u}x \in \mathrm{Prf}(u) \cap \omega(wu^R\bar{v})$, where $x \in \mathrm{A}$. Suppose that there is $\bar{u}x \in \mathrm{Prf}(u) \cap \omega(wu^R\bar{v})$. Let $y \in \mathrm{A}$ be such that $x \neq y$ and $wu^R\bar{v}\bar{u}y \in \mathrm{R}$. Let $t = \mathrm{lps}(wu^R\bar{v}\bar{u}y)$. We distinguish two cases:

- $t \in \mathrm{A}$. The assumptions of the proposition guarantee that $t \in \mathrm{F}(w)$.
- $t = y\bar{t}y$ for some palindrome $\bar{t}$. Clearly $y\bar{t}x \in \mathrm{swSuf}(wu^R\bar{v}, u)$ and the assumptions of the proposition guarantee that $t = \mathrm{spc}(ytx) = yty \in \mathrm{F}(w)$.

It follows that the longest palindromic suffix t is not unioccurrent, hence $wu^R\bar{v}\bar{u}y$ is not rich; see Proposition 2.3. This completes the proof. $\square$

Given a factor u of a word w, for us it will not be important if u or u^R is unioccurrent in w. For this purpose we define a special notion.

Definition 5.5. *If* $\sum_{v\in\{u,u^R\}} \mathrm{occur}(w,v) = 1$ *then we say that the word* u *is* reverse-unioccurrent *in* w, *where* $w, u \in \mathrm{A}^+$.

Remark 5.6. The notion of reverse-unioccurrence has also been used in [4].

We show that if the switch ytx is a suffix of the word wx and ytx is reverse-unioccurrent in wx then wx is a flexed point of wx.

Lemma 5.7. *If* $w, wx \in \mathrm{R}$*,* $x, y \in \mathrm{A}$*,* $ytx \in \mathrm{Suf}(wx) \cap \mathrm{sw}(wx)$*, and* ytx *is reverse-unioccurrent in* wx *then* $wx \in \mathrm{T}(wx)$*.*

Proof. Suppose that $wx \in \mathrm{ER_a}(w)$. If $u = \mathrm{lpps}(w)$ then $|t| < |u|$ and $t \in \mathrm{Prf}(u) \cap \mathrm{Suf}(u)$. It follows that $xux \in \mathrm{Suf}(wx)$ and $ty \in \mathrm{Prf}(u)$, since $yt \in \mathrm{Suf}(u)$. Consequently $xty \in \mathrm{Prf}(xu)$, which is a contradiction, because xty is reverse-unioccurrent in wx. The lemma follows. □

There is an obvious corollary of Lemma 5.7 saying that if t is a switch of w, then there is a flexed point v of w such that either t or t^R is a suffix of v.

Corollary 5.8. *If* $w \in \mathrm{R}$*,* $t \in \mathrm{sw}(w)$ *then there is* $v \in \mathrm{T}(w)$ *such that* $\{t, t^R\} \cap \mathrm{Suf}(v) \neq \emptyset$*.*

Proof. If $w \in \mathrm{R}$ and $t \in \mathrm{sw}(w)$, then there is obviously $u \in \mathrm{Prf}(w)$ such that $\{t, t^R\} \cap \mathrm{Suf}(u) \neq \emptyset$ and t is reverse-unioccurrent in u. Then Lemma 5.7 implies that $u \notin \mathrm{ER_a}(\mathrm{rtrim}(u))$. This completes the proof. □

In order to construct a word with a prefix containing all switch palindromic closures of its switches we introduce two functions ewp and elpp.

Definition 5.9. *If* $w, t \in \mathrm{R} \cap \mathrm{A}^+$ *and* t *is a palindrome then we define*

$$\Sigma_{w,t} = \{u \mid u \in \mathrm{Prf}(w) \text{ and } |u| \geq |\,\mathrm{lppp}(w)| \text{ and } \mathrm{rtrim}(t) \in \mathrm{Suf}(u)\}.$$

If $\Sigma_{w,t} \neq \emptyset$ *then let* $\bar{\pi}_{w,t}$ *denote the shortest element of* $\Sigma_{w,t}$ *and let* $\pi_{w,t}$ *be such that* $\bar{\pi}_{w,t} = \mathrm{lppp}(w)\pi_{w,t}$*.*

Let $x = \mathrm{Prf}(t) \cap \mathrm{A}$ *and let*

$$\mathrm{ewp}(w,t) = \begin{cases} x(\pi_{w,t})^R w & \text{if } \Sigma_{w,t} \neq \emptyset \text{ and } t \notin \mathrm{F}(v^R w) \\ w & \text{otherwise.} \end{cases}$$

In addition we define

$$\mathrm{ewp}(w, t_1, t_2, \ldots, t_m) = \mathrm{ewp}(\ldots(\mathrm{ewp}(\mathrm{ewp}(w, t_1), t_2), \ldots), t_m),$$

where w *is a nonempty rich word and* $t_1, t_2, \ldots, t_m$ *are rich nonempty palindromes.*

Given $w \in \mathrm{A}^+$ *and* $x \in \mathrm{A}$*, let* $\mathrm{maxPow}(w, x) = k$ *such that* $x^k \in \mathrm{F}(w)$ *and* $x^{k+1} \notin \mathrm{F}(w)$*.*

Suppose $w \in \mathrm{R}$*,* $y \in \mathrm{A}$*, and* $k = \mathrm{maxPow}(w, y)$*. Let* $\mathrm{elpp}_y(w) = \mathrm{ewp}(w, y^{k+1})$*.*

Remark 5.10. The notation "ewp" stands for "extension with prefix". It is clear that $(\pi_{w,t})^R w$ is a left standard extension of w that has as a prefix $\mathrm{ltrim}(t)$.

The notation "maxPow" stands for "maximal power". If $x \notin \mathrm{F}(w)$ then $\mathrm{maxPow}(w, x) = 0$.

The notation "elpp" stands for "extension with letter power prefix". The function $\mathrm{elpp}_y(w)$ is the word yu where u is a left standard extension of w such that $y^{\mathrm{maxPow}(w,y)}$ is a prefix. If $\mathrm{maxPow}(w, y) = 0$ then $\mathrm{elpp}_y(w) = yw$.

Example 5.11. Let $\mathrm{A} = \{0, 1, 2\}$, $w = 2020111010111010$, $t_1 = 11011$, and $t_2 = 20201$. Then we have:

- $\mathrm{rtrim}(t_1) = 1101$, $\mathrm{ltrim}(t_1) = 1011$, $\mathrm{lpp}(w) = 202$.
- $\Sigma_{w,t_1} = \{202011101, 202011101011101\}$, $\sigma_{w,t_1} = 011101$.
- $\mathrm{ewp}(w, t_1) = 1101110202011101011101 0$.
- Let $v = \mathrm{ewp}(w, t_1)$. Then $\sigma_{v,t_2} = 102020$
- $\mathrm{ewp}(w, t_1, t_2) = \mathrm{ewp}(v, t_2) = 20202011101110202011101011101 0$.
- $\mathrm{maxPow}(w, 1) = 3$, $\mathrm{maxPow}(w, 2) = 1$, and $\mathrm{maxPow}(w, 0) = 1$.
- $\mathrm{elpp}_1(w) = \mathrm{ewp}(w, 1111) = 1111020201110101110 10$.
- $\mathrm{elpp}_2(w) = \mathrm{ewp}(w, 22) = 22020111010111010$.
- $\mathrm{elpp}_0(w) = \mathrm{ewp}(w, 00) = 002020111010111010$.

We prove that $\mathrm{ewp}(w, t), \mathrm{elpp}_y(w) \in \mathrm{R}$ are rich words.

Lemma 5.12. *If* $w, t \in \mathrm{R} \cap \mathrm{A}^+$ *and* $y \in \mathrm{A}$ *then* $\mathrm{ewp}(w, t), \mathrm{elpp}_y(w) \in \mathrm{R}$.

Proof. Because $\mathrm{elpp}(w) = \mathrm{ewp}_y(w, y^{k+1})$ it suffices to prove that $\mathrm{ewp}(w, t) \in \mathrm{R}$. From the definition of $\mathrm{ewp}(w, t)$ it is clear that we need to verify only the case where $\Sigma_{w,t} \neq \emptyset$ and $t \notin \mathrm{F}(v^R w)$. Obviously $(\pi_{w,t})^R w \in \mathrm{R}$, since $(\pi_{w,t})^R w \in \mathrm{EL_a}(w)$, see Lemma 3.4. Let $x = \mathrm{Prf}(t) \cap \mathrm{A}$. Then $\mathrm{lpp}(xv^R w) = t$ and since $t \notin \mathrm{F}(v^R w)$ we have $\mathrm{occur}(xv^R w, t) = 1$. Hence Corollary 2.4 implies that $xv^R w \in \mathrm{R}$. $\square$

6 Construction of a Uniquely Extensible Rich Word II

In this section we consider that $\{0,1\} \subseteq \mathrm{A}$. Let $\mathrm{g}_n = \mathrm{g}_{n-1}\, 01^n 0\, \mathrm{g}_{n-1}$, where $\mathrm{g}_1 = 1$ and $n > 1$. For $n, k \geq 2$ we show that the words $0^k\, \mathrm{g}_n$ are rich and that $0^k\, \mathrm{g}_{n-1}\, 01, 0^k\, \mathrm{g}_{n-1}\, 01^n$ are the only flexed points of $0^k\, \mathrm{g}_n$ that are not flexed points of $0^k\, \mathrm{g}_{n-1}$. Let $\bar{\mathrm{T}}_n = \mathrm{T}(0^k\, \mathrm{g}_n) \setminus \mathrm{T}(0^k\, \mathrm{g}_{n-1})$.

Proposition 6.1. *If $n, k \geq 2$ then $0^k g_n \in \mathrm{R}$ and*

$$\bar{\mathrm{T}}_n = \{0^k\, \mathrm{g}_{n-1}\, 01, 0^k\, \mathrm{g}_{n-1}\, 01^n\}.$$

Proof. Obviously $0^k\, \mathrm{g}_1 \in \mathrm{R}$. Suppose that $0^k\, \mathrm{g}_{n-1} \in \mathrm{R}$, where $n \geq 2$. We show that $0^k\, \mathrm{g}_n \in \mathrm{R}$. We have that $0^k\, \mathrm{g}_n = 0^k\, \mathrm{g}_{n-1}\, 01^n 0\, \mathrm{g}_{n-1}$. Note that $\mathrm{lps}(0^k\, \mathrm{g}_{n-1}) = \mathrm{g}_{n-1}$. It follows that $0^k\, \mathrm{g}_{n-1}\, 0 = \mathrm{ER}(0^k\, \mathrm{g}_{n-1})$ and hence $0^k\, \mathrm{g}_{n-1}\, 0 \in \mathrm{R}$. It is easy to see that

$$\mathrm{lps}(0^k\, \mathrm{g}_{n-1}\, 01) = \mathrm{lps}(0^k\, \mathrm{g}_{n-2}\, 01^{n-1} 0\, \mathrm{g}_{n-2}\, 01) = 10\, \mathrm{g}_{n-2}\, 01$$

and that $\mathrm{occur}(0^k\, \mathrm{g}_{n-1}\, 01, 10\, \mathrm{g}_{n-2}\, 01) = 1$. Hence we have $0^k\, \mathrm{g}_{n-1}\, 01 \in \mathrm{R}$; see Proposition 2.3. It follows that $0^k\, \mathrm{g}_{n-1}\, 01^{n-1} \in \mathrm{ER_a}(0^k\, \mathrm{g}_{n-1}\, 01) \subseteq \mathrm{R}$. Also we have that $0^k\, \mathrm{g}_{n-1}\, 01 \neq \mathrm{ER}(0^k\, \mathrm{g}_{n-1}\, 0)$ and thus $0^k\, \mathrm{g}_{n-1}\, 01 \in \bar{\mathrm{T}}_n$.

Obviously $\mathrm{occur}(0^k\, \mathrm{g}_{n-1}\, 01^n, 1^n) = 1$. Since 1^n is a palindrome we have that $0^k\, \mathrm{g}_{n-1}\, 01^n \in \mathrm{R}$; see Proposition 2.3. Since $\mathrm{g}_{n-1}\, 01^n 0\, \mathrm{g}_{n-1}$ is a palindrome we have that $\mathrm{lps}(0^k\, \mathrm{g}_{n-1}\, 01^n t) = t^R 1^n t$ for each $t \in \mathrm{Prf}(0\, \mathrm{g}_{n-1})$. This implies that $0^k\, \mathrm{g}_{n-1}\, 01^n t \in \mathrm{ER_a}(0^k\, \mathrm{g}_{n-1}\, 01^n) \subseteq \mathrm{R}$ and in particular $0^k\, \mathrm{g}_n \in \mathrm{ER_a}(0^k\, \mathrm{g}_{n-1}\, 01^n) \subseteq \mathrm{R}$. Clearly $0^k\, \mathrm{g}_{n-1}\, 01^n \neq \mathrm{ER}(0^k\, \mathrm{g}_{n-1}\, 01^{n-1})$ and thus $0^k\, \mathrm{g}_{n-1}\, 01^n \in \bar{\mathrm{T}}_n$.

Consequently for each $n, k \geq 2$, we conclude that $0^k\, \mathrm{g}_n \in \mathrm{R}$ and $\bar{\mathrm{T}}_n = \{0^k\, \mathrm{g}_{n-1}\, 01, 0^k\, \mathrm{g}_{n-1}\, 01^n\}$. □

We present all switches of $0^k\, \mathrm{g}_n$. Let $\mathrm{S}_n = \big(\mathrm{sw}(0^k\, \mathrm{g}_n) \setminus \mathrm{sw}(0^k\, \mathrm{g}_{n-1})\big) \cap \bigcup_{w \in \bar{\mathrm{T}}_n} \mathrm{Suf}(w)$, where $n \geq 3$.

Proposition 6.2. *If $k \geq 2$ and $n \geq 3$ then*

$$\mathrm{S}_n = \{00\, \mathrm{g}_{n-1}\, 01, 01^{n-1} 0\, \mathrm{g}_{n-2}\, 01^n, 01^n\}.$$

Proof. Proposition 6.1 states that $\bar{\mathrm{T}}_n = \{0^k\, \mathrm{g}_{n-1}\, 01, 0^k\, \mathrm{g}_{n-1}\, 01^n\}$. We will consider the switches that are suffixes of the flexed points from $\bar{\mathrm{T}}_n$:

- For $0^k \mathrm{g}_{n-1} 01$: Let $t = \mathrm{lps}(0^k \mathrm{g}_{n-1} 0)$. Obviously $t = 0 \mathrm{g}_{n-1} 0$. Since $\mathrm{occur}(t, 1^{n-1}) = 1$ it follows that t is the only palindromic suffix of $0^k \mathrm{g}_{n-1} 0$ which contains the factor 1^{n-1}. Consequently each palindromic suffix of $0^k \mathrm{g}_{n-1} 0$ which is not equal to t is a factor of $0 \mathrm{g}_{n-2} 0 \in \mathrm{Suf}(t)$. Thus $00 \mathrm{g}_{n-1} 01$ is the only switch of $\bar{t} = 00 \mathrm{g}_{n-1} 01$ which is not a switch of $0 \mathrm{g}_{n-2} 0 \in \mathrm{F}(0^k \mathrm{g}_{n-1})$.
- For $0^k \mathrm{g}_{n-1} 01^n$: Let $t \in \mathrm{Suf}(0^k \mathrm{g}_{n-1} 01^n) \cap \mathrm{sw}(0^k \mathrm{g}_n)$. Since $1^n \in \mathrm{Suf}(00 \mathrm{g}_{n-1} 01^n)$ it follows that $|t| \geq n+1$. For $|t| = n+1$ there is the switch 01^n. For $|t| > n+1$ we have that $1^{n-1} \in \mathrm{Suf}(\mathrm{rtrim}(t)) \cap \mathrm{Prf}(\mathrm{ltrim}(t))$ and because $\mathrm{occur}(00 \mathrm{g}_{n-1} 01^{n-1}, 1^{n-1}) = 2$ it follows that there is only one switch with $|t| > n+1$; namely $\bar{t} = 01^{n-1}0 \mathrm{g}_{n-2} 01^n$.

The proposition follows. □

Proposition 6.2 and Corollary 5.8 allow us to list all switches of $0^k \mathrm{g}_n$.

Corollary 6.3. *If $n \geq 3$ then*

$$\mathrm{sw}(0^k \mathrm{g}_n) = \bigcup_{i=1}^{k} \{00^i1\} \cup \{01, 10, 00101, 11010, 01011\} \cup$$
$$\bigcup_{i=3}^{n} \{00 \mathrm{g}_{n-1} 01, 01^{n-1}0 \mathrm{g}_{n-2} 01^n, 1^n 0 \mathrm{g}_{n-2} 01^{n-1}0, 01^n, 1^n 0\}.$$

Proof. Proposition 6.2 states that $\mathrm{S}_n = \{00 \mathrm{g}_{n-1} 01, 01^{n-1}0 \mathrm{g}_{n-2} 01^n, 01^n\}$ for $n \geq 3$. We may easily check that

- $(00 \mathrm{g}_{n-1} 01)^R = 10 \mathrm{g}_{n-1} 00 \notin \mathrm{F}(0^k \mathrm{g}_n)$,
- $(01^{n-1}0 \mathrm{g}_{n-2} 01^n)^R = 1^n 0 \mathrm{g}_{n-2} 01^{n-1}0 \in \mathrm{F}(0^k \mathrm{g}_n)$, and
- $(00 \mathrm{g}_{n-1} 01)^R \notin \mathrm{F}(0^k \mathrm{g}_m)$ for all $m \geq 2$.

Obviously $\mathrm{sw}(0^k \mathrm{g}_2) = \bigcup_{i=1}^{k} \{00^i1\} \cup \{01, 10, 001, 00101, 01011\}$; recall that $0^k \mathrm{g}_2 = 0^k 101101$. Note that $(01011)^R = 11010 \in \mathrm{F}(\mathrm{g}_3)$, $(00^i1)^R = 10^i0 \notin \mathrm{F}(\mathrm{g}_m)$, and $(00101)^R = 10100 \notin \mathrm{F}(\mathrm{g}_m)$ for all $i, m \geq 1$. Corollary 5.8 asserts for every switch t of w that there is a flexed points $\bar{w} \in \mathrm{T}(w)$ such that $\{t, t^R\} \cap \mathrm{Suf}(\bar{w}) \neq \emptyset$. The corollary follows. □

Let $j \geq 2$. We define:

- $\alpha_{1,j} = 00\,\mathrm{g}_{j-1}\,00$,
- $\alpha_{2,j} = 01^{j-1}0\,\mathrm{g}_{j-2}\,01^{j-1}0$,
- $\alpha_{3,j} = 1^{j}0\,\mathrm{g}_{j-2}\,01^{j}$, and
- $\alpha_{4,j} = 1^{j+1}$.

The next obvious corollary of Corollary 6.3 presents the switch palindromic closures of all switches of the word $0^k\,\mathrm{g}_n$.

Corollary 6.4. *If $k \geq 2$ and $n \geq 3$ then*

$$\mathrm{spc}(\mathrm{sw}(0^k\,\mathrm{g}_n)) = \{0^k, 00100, 11011, 01010, \alpha_{4,n}\} \cup \bigcup_{i=3}^{n} \{\alpha_{1,j}, \alpha_{2,j}, \alpha_{3,j}\}.$$

Proof. Corollary 6.3 lists all switches of the word $0^k\,\mathrm{g}_n$. For every switch $t \in \mathrm{sw}(0^k\,\mathrm{g}_n)$ we may easily verify that there is $v \in \mathrm{spc}(\mathrm{sw}(0^k\,\mathrm{g}_n))$ such that $\mathrm{spc}(t) \in \mathrm{F}(v)$. This completes the proof. $\square$

Remark 6.5. Note that the palindromes $\alpha_{4,j}$ are factors of $\alpha_{4,n}$ for $j \leq n$. This is the difference to palindromes $\alpha_{i,j}$, where $i \in \{1, 2, 3\}$. For this reason the palindrome $\alpha_{4,j}$ is not involved in the union formula from $i = 3$ to n.

The next definition defines a word h_n. Later we show that h_n is a unique rich extension of $\bar{\mathrm{h}}_n$, where $\mathrm{h}_n = \bar{\mathrm{h}}_n\,\mathrm{ltrim}(\mathrm{g}_n)$.

Definition 6.6. *Let $n \geq 3$. We define:*

- $\kappa(j, w) = \mathrm{elpp}_0(\mathrm{ewp}(w, \alpha_{1,j}, \alpha_{2,j}, \alpha_{3,j}, \alpha_{4,j}))$, *where $w \in \mathrm{R} \cap \mathrm{A}^+$ and $3 \leq j \leq n$.*
- $\mathrm{h}_{n,n} = \kappa(n, 000\,\mathrm{g}_n\,00\,\mathrm{g}_n)$.
- $\mathrm{h}_{n,j} = \kappa(j, \mathrm{h}_{n,j+1})$, *where $3 \leq j < n$.*
- *Suppose that* A *is totally ordered, let $\sigma(\mathrm{A}) = x_1 x_2 \ldots x_m$, where $x_i \in \mathrm{A} \setminus \{0, 1\}$, $x_i < x_{i+1}$, $1 \leq i < m$, and $m = |\mathrm{A}| - 2$.*
- $\mathrm{h}_n = \sigma(\mathrm{A})\,\mathrm{ewp}(\mathrm{h}_{n,3}, 00100, 11011, 01010)$.

Remark 6.7. The function $\kappa(j,w)$ extends the word w to a word $\bar{w}w$ in such a way that $\bar{w}w$ contains the switch palindromic closures of switches $\alpha_{1,j}$, $\alpha_{2,j}$, $\alpha_{3,j}$, $\alpha_{4,j}$. In addition the longest palindromic prefix of $\bar{w}w$ is 0^k for some $k > 0$.

The word $\mathrm{h}_{n,3}$ is constructed by iterative applying of the function $\kappa(j,w)$ starting with the word $000\,\mathrm{g}_n\,00\,\mathrm{g}_n$.

The word h_n contains the switch palindromic closures of all switches of the word $00\,\mathrm{g}_n$. The suffix of h_n is the word $000\,\mathrm{g}_n\,00\,\mathrm{g}_n$. As a result h_n has the form $u\,\mathrm{rtrim}(\mathrm{g}_n)1001\,\mathrm{ltrim}(\mathrm{g}_n)$ for some $u \in \mathrm{A}^*$. It is the form used in Proposition 5.4. The prefix $\sigma(\mathrm{A})$ of h_n is there to assert that u contains all letters. The order of the letters does not matter.

We show that h_n *is a rich word.*

Lemma 6.8. *If* $n \geq 3$ *then* $\mathrm{h}_n \in \mathrm{R}$.

Proof. Lemma 5.12 says that both $\mathrm{ewp}(w,t), \mathrm{elpp}_0(w) \in \mathrm{R}$, where $w,t \in \mathrm{R}$. Proposition 2.2 and Proposition 6.1 imply that $\mathrm{rtrim}(\alpha_{i,j}) \in \mathrm{R}$, since $\mathrm{rtrim}(\alpha_{i,j}) \in \mathrm{F}(00\,\mathrm{g}_j) \subseteq \mathrm{F}(00\,\mathrm{g}_n)$, where $i \in \{1,2,3,4\}$ and $3 \leq j \leq n$. Because $\alpha_{i,n} = \mathrm{ER}(\mathrm{rtrim}(\alpha_{i,n}))$ we have that $\alpha_{i,n} \in \mathrm{R}$, see Lemma 3.4. Hence $\kappa(j,w) \in \mathrm{R}$.

Proposition 6.1 asserts that $0^k\,\mathrm{g}_n \in \mathrm{R}$. Also it is easy to see that $000\,\mathrm{g}_n\,00\,\mathrm{g}_n \in \mathrm{R}$; just consider that $00\,\mathrm{g}_n\,00\,\mathrm{g}_n \in \mathrm{ER}_{\mathrm{a}}(00\,\mathrm{g}_n)$,

$$\mathrm{occur}(000\,\mathrm{g}_n\,00\,\mathrm{g}_n, 000) = 1, \text{ and } \mathrm{lpp}(000\,\mathrm{g}_n\,00\,\mathrm{g}_n) = 000,$$

see Corollary 2.4. In consequence $\mathrm{h}_{n,j} \in \mathrm{R}$ for $3 \leq j < n$. We have that $\mathrm{ewp}(\mathrm{h}_{n,3}, 00100, 11011, 01010) \in \mathrm{R}$, because $00100, 01010, 11011 \in \mathrm{R}$.

Obviously $\sigma(\mathrm{A}) \in \mathrm{R}$. Moreover it is easy to verify that if $w_1, w_2 \in \mathrm{R}$ and $\mathrm{F}(w_1) \cap \mathrm{F}(w_2) = \epsilon$ then $w_1w_2 \in \mathrm{R}$. Hence

$$\sigma(\mathrm{A})\,\mathrm{ewp}(\mathrm{h}_{n,3}, 00100, 11011, 01010) \in \mathrm{R}\,.$$

We conclude that $\mathrm{h}_n \in \mathrm{R}$. □

Proposition 6.9. *Let* $\bar{\mathrm{h}}_n$ *be such that* $\mathrm{h}_n = \bar{\mathrm{h}}_n\,\mathrm{ltrim}(\mathrm{g}_n)$. *If* $n > 2$ *then* h_n *is a unique rich extension of* $\bar{\mathrm{h}}_n$.

Proof. Obviously there is $w \in \mathrm{R}$ such that $\mathrm{h}_n = w\,\mathrm{rtrim}(\mathrm{g}_n)1001\,\mathrm{ltrim}(\mathrm{g}_n)$. Corollary 6.4 lists the elements of $\mathrm{spc}(\mathrm{sw}(0^k\,\mathrm{g}_n))$. The construction of h_n guarantees that all these elements are factors of w; formally $\alpha_{i,j} \in \mathrm{F}(w)$ for all $i \in \{1,2,3,4\}$ and $3 \leq j \leq n$. For $00\,\mathrm{g}_2\,0 = 001011010$ we can see that $\mathrm{sw}(001011010) = \{01, 10, 001, 011, 110, 00101, 01011, 11010\}$. It follows that $\mathrm{spc}(\mathrm{sw}(001011010)) = \{00100, 01010, 000, 111, 11011\}$. Obviously we have that $\{00100, 01010, 000, 111, 11011\} \subseteq \mathrm{F}(w)$.

Since $\sigma(\mathrm{A}) \in \mathrm{Prf}(\mathrm{h}_n)$ we have that $\mathrm{A} \in \mathrm{F}(w)$. It is easy to check that $\mathrm{lps}(w\,\mathrm{rtrim}(\mathrm{g}_n)1001) = 1001$. Hence we have

$$w\,\mathrm{rtrim}(\mathrm{g}_n)1001\,\mathrm{ltrim}(\mathrm{g}_n) \in \mathrm{ER}_\mathrm{a}(w\,\mathrm{rtrim}(\mathrm{g}_n)1001).$$

Thus Proposition 5.4 implies that h_n is a unique rich extension of $\bar{\mathrm{h}}_n$. $\square$

Let $\rho(n) = |\,\mathrm{g}_n\,|$, where $n \geq 1$. Since $\mathrm{g}_n = \mathrm{g}_{n-1}\,01^n0\,\mathrm{g}_{n-1}$, we have $2\rho(n) < \rho(n+1)$ and consequently $\rho(n) < \frac{1}{2^k}\rho(n+k)$, where $k > 0$.

We derive an upper bound for length of h_n. We start with an upper bound for $|\kappa(j,w)|$.

Proposition 6.10. *If $j,k > 2$, $\bar{w} \in \mathrm{R}$, $w = 0^k\,\mathrm{g}_j\,\bar{w} \in \mathrm{R}$, $\mathrm{lpp}(w) = 0^k$ and $\alpha_{i,j} \notin \mathrm{F}(\bar{w})$ for $i \in \{1,2,3\}$ then $|\kappa(j,w)| < |w| + 7\rho(j-1) + 5k + 5j + 10$.*

Proof. Let $t_1 = \mathrm{ewp}(w, \alpha_{1,j})$, $t_2 = \mathrm{ewp}(t_1, \alpha_{2,j})$, $t_3 = \mathrm{ewp}(t_2, \alpha_{3,j})$, and $t_4 = \mathrm{ewp}(t_3, \alpha_{4,j})$. Clearly $\kappa(j,w) = \mathrm{elpp}_0(t_4)$. It is easy to see that:

- $t_1 = 00\,\mathrm{g}_{j-1}\,0^k\,\mathrm{g}_j\,\bar{w}$; $\mathrm{lpp}(t_1) = \alpha_{1,j} = 00\,\mathrm{g}_{j-1}\,00$.
- $t_2 = 01^{j-1}0\,\mathrm{g}_{j-2}\,01^{j-1}0\,\mathrm{g}_{j-2}\,0^{k-2}t_1$;
  $\mathrm{lpp}(t_2) = \alpha_{2,j} = 01^{j-1}0\,\mathrm{g}_{j-2}\,01^{j-1}0$.
- $t_3 = 1^j0\,\mathrm{g}_{j-2}\,01^j0\,\mathrm{g}_{j-1}\,0^k\,\mathrm{g}_{j-1}\,0^k\,\mathrm{g}_{j-2}\,t_2$; $\mathrm{lpp}(t_3) = \alpha_{3,j} = 1^j0\,\mathrm{g}_{j-2}\,01^j$.
- If $\alpha_{4,j} \in \mathrm{F}(w)$ then $t_4 = t_3$ and $\mathrm{lpp}(t_4) = \mathrm{lpp}(t_3)$ else $t_4 = 1t_3$ and $\mathrm{lpp}(t_4) = \alpha_{4,j} = 1^{j+1}$.
- If $\alpha_{4,j} \in \mathrm{F}(w)$ then $\kappa(j,w) = 00^k\,\mathrm{g}_{j-1}\,0t_4$ else

  $$\kappa(j,w) = 00^k\,\mathrm{g}_{j-1}\,01^j0\,\mathrm{g}_{j-2}\,0t_4.$$

  In either case we have $\mathrm{lpp}(\kappa(j,w)) = 0^{k+1}$.

It follows that:

- $|t_1| = |w| + \rho(j-1) + 2$.

- $|t_2| = |t_1| + (k-2) + 2\rho(j-2) + 2(j-1) + 4 = |w| + 6 + \rho(j-1) + 2\rho(j-2) + (k-2) + 2(j-1)$.
- $|t_3| = |t_2| + 2\rho(j-2) + 2\rho(j-1) + 2k + 3 + 2j = |w| + 9 + 3\rho(j-1) + 4\rho(j-2) + (k-2) + 2k + 2(j-1) + 2j$.
- $|t_4| \leq |t_3| + 1$.
- $|\kappa(j,w)| \leq |t_4| + k + 4 + \rho(j-1) + \rho(j-2) + j = |w| + 14 + 4\rho(j-1) + 5\rho(j-2) + (k-2) + 3k + 2(j-1) + 3j$.

Since $2\rho(j-2) < \rho(j-1)$ we have $|\kappa(j,w)| < |w| + 7\rho(j-1) + 4k + 5j + 10$. $\square$

The main theorem of the section presents an upper bound for the length of h_n.

Theorem 6.11. *If* $n \geq 2$ *then* $|\,\mathrm{h}_n\,| < \frac{11}{2}\rho(n) + (n-3)(5n+22) + 3n + 20 + |\mathrm{A}|$.

Proof. Proposition 6.10 implies for $j = n$, $k = 3$ and $w = 000\,\mathrm{g}_n\,00\,\mathrm{g}_n$ that

$$|\,\mathrm{h}_{n,n}\,| = |\kappa(n,w)| < |w| + 7\rho(n-1) + 4*3 + 5n + 10. \tag{1}$$

For $n-1$ and $n-2$ we have:

- $|\,\mathrm{h}_{n,n-1}\,| = |\kappa(n-1, \mathrm{h}_{n,n})| < |\,\mathrm{h}_{n,n}\,| + 7\rho(n-2) + 4*4 + 5(n-1) + 10$.
- $|\,\mathrm{h}_{n,n-2}\,| = |\kappa(n-2), \mathrm{h}_{n,n-1}\,| < |\,\mathrm{h}_{n,n-1}\,| + 7\rho(n-3) + 4*5 + 5(n-2) + 10$.

And generally for $n-i$:

$$\begin{aligned}|\,\mathrm{h}_{n,n-i}\,| < |\,\mathrm{h}_{n,n-i+1}\,| + 7\rho(n-i-1) + 4(i+3) + 5(n-i) + 10 = \\ |\,\mathrm{h}_{n,n-i+1}\,| + 7\rho(n-i-1) + 5n - i + 22 < \\ |\,\mathrm{h}_{n,n-i+1}\,| + 7\rho(n-i-1) + 5n + 22.\end{aligned} \tag{2}$$

Realize that $\rho(n-i-1) \leq \frac{1}{2^{i+1}}\rho(n)$, $|w| = 2\rho(n) + 5$, and $\sum_{i=1}^{n-3} \frac{1}{2^{i+1}} < \frac{1}{2}$. It follows from (1) and (2) that:

$$\begin{aligned}|\,\mathrm{h}_{n,3}\,| < 2\rho(n) + 5 + 7\rho(n)\sum_{i=1}^{n-3}\frac{1}{2^{i+1}} + (5n+22)\sum_{i=1}^{n-3} 1 < \\ \frac{11}{2}\rho(n) + (n-3)(5n+22) + 5.\end{aligned} \tag{3}$$

Obviously $\mathrm{lpp}(\mathrm{h}_{n,3}) = 0^n$ and $0^n\,\mathrm{g}_2\,0 \in \mathrm{Prf}(\mathrm{h}_{n,3})$.

Let $t_1 = \mathrm{ewp}(\mathrm{h}_{n,3}, 00100)$, $t_2 = \mathrm{ewp}(t_1, 11011)$. We have that $\mathrm{h}_n = \sigma(\mathrm{A})\,\mathrm{ewp}(t_2, 01010)$. We can verify that $t_1 = 001\,\mathrm{h}_{n,3}$, $t_2 = 11011010^{n-2}t_1$, and $\mathrm{h}_n = \sigma(\mathrm{A})01010^n10^n10t_2$.

Using (3) we get:

$$|\,\mathrm{h}_n\,| \leq |\,\mathrm{h}_{n,3}\,| + |001| + |11011010^{n-2}| + |01010^n10^n10| + |\mathrm{A}| < \frac{11}{2}\rho(n) + (n-3)(5n+22) + 3n + 20 + |\mathrm{A}|\,.$$

This completes the proof. ☐

Theorem 6.11 and Proposition 6.9 have the following corollary to the lower bound for $\phi(n)$.

Corollary 6.12. *For each real constant $c > 0$ and each integer $m > 0$ there is $n > m$ such that $\phi(n) \geq (\frac{2}{9} - c)n$.*

Proof. Proposition 6.9 implies that $\omega(|\bar{\mathrm{h}}_n|) \leq |\,\mathrm{g}_n\,| - 1 = \rho(n) - 1$. It follows that $\phi(|\bar{\mathrm{h}}_n|) \geq \rho(n) - 1$ and

$$\phi(|\bar{\mathrm{h}}_n|) \geq \frac{\rho(n)-1}{|\bar{\mathrm{h}}_n|}|\bar{\mathrm{h}}_n|. \tag{4}$$

From Theorem 6.11 and Proposition 6.9 we have that

$$\frac{\rho(n)-1}{|\bar{\mathrm{h}}_n|} = \frac{\rho(n)-1}{|\,\mathrm{h}_n\,| - \rho(n) - 1} = \frac{\rho(n)-1}{\frac{9}{2}\rho(n) + (n-3)(5n+22) + |\mathrm{A}| + 4}.$$

Since $\rho(n) \geq 2^n$ this implies that

$$\frac{\rho(n)-1}{|\bar{\mathrm{h}}_n|} \leq \frac{2}{9} \text{ for } n > 3 \tag{5}$$

and

$$\lim_{n\to\infty} \frac{\rho(n)-1}{|\bar{\mathrm{h}}_n|} = \frac{2}{9} \tag{6}$$

The corollary follows from (4),(5), and (6). ☐

Acknowledgments

The author acknowledges support by the Czech Science Foundation grant GAČR 13-03538S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/205/OHK4/3T/14.

References

[1] 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.

[2] 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.

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

[4] J. Rukavicka, *Rich words containing two given factors*, Mercaş R., Reidenbach D. (eds) Combinatorics on Words. WORDS 2019. Lecture Notes in Computer Science, vol 11682. Springer, Cham, DOI: https://doi.org/10.1007/978-3-030-28796-2_23.

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

Article [[Ru06]]: Upper bound for the number of closed and privileged words

Information Processing Letters 156 (2020) 105917

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Upper bound for the number of closed and privileged words

Josef Rukavicka

Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic

A R T I C L E   I N F O

Article history:
Received 26 November 2019
Received in revised form 13 January 2020
Accepted 14 January 2020
Available online 16 January 2020
Communicated by Marek Chrobak

Keywords:
Closed words
Privileged words
Upper bound
Combinatorial problems

A B S T R A C T

A non-empty word w is a *border* of the word u if $|w| < |u|$ and w is both a prefix and a suffix of u. A word u with the border w is *closed* if u has exactly two occurrences of w. A word u is *privileged* if $|u| \leq 1$ or if u contains a privileged border w that appears exactly twice in u.

Peltomäki (2016) presented the following open problem: "Give a nontrivial upper bound for $B(n)$", where $B(n)$ denotes the number of privileged words of length n. Let $\mathrm{D}(n)$ denote the number of closed words of length n. Let $q > 1$ be the size of the alphabet. We show that there is a positive real constant c such that

$$\mathrm{D}(n) \leq c \ln n \frac{q^n}{\sqrt{n}}, \text{ where } n > 1.$$

Privileged words are a subset of closed words, hence we show also an upper bound for the number of privileged words.

1. Introduction

A non-empty word w is a *border* of the word u if $|w| < |u|$ and w is both a prefix and a suffix of u. A border w of the word u is the *maximal border* of u if for every border $\bar{w}$ of u we have that $|\bar{w}| \leq |w|$. A word u with the border w is *closed* if u has exactly two occurrences of w. It follows that w occurs only as a prefix and as a suffix of u. A word u is *privileged* if $|u| \leq 1$ or if u contains a privileged border w that appears exactly twice in u. Obviously privileged words are a subset of closed words.

The properties of closed and privileged words have been studied in recent years [2], [5], [6]. One of the questions that has been investigated is the enumeration of privileged words. In [3], it was proved that there are constants c and n_0 such that for all $n > n_0$, there are at least $\frac{cq^n}{n(\log_q n)^2}$ privileged words of length n. This improves the lower bound for the number of privileged words from [1]. Since every privileged word is a closed word, the result from [3] forms also a lower bound for the number of closed words.

Concerning an upper bound for the number of privileged words we have found only the following open problem [4]: "Give a nontrivial upper bound for $B(n)$", where $B(n)$ denotes the number of privileged words of length n. Also in [4], the author presents an idea how to improve the lower bound from [3]. On the other hand, in [4], there is no explicit suggestion how to approach the problem of determining the upper bound.

In the current article we construct an upper bound for the number of closed words of length n. Since the privileged words are a subset of closed words, we present also a response to the open problem from [4].

E-mail address: josef.rukavicka@seznam.cz.

https://doi.org/10.1016/j.ipl.2020.105917

We explain in outline our proof. Let A be an alphabet with $q>1$ letters, let A^m denote the set of all words of length m, and let $\mathrm{A}^*=\bigcup_{m\geq 0}\mathrm{A}^m$. It is known that $|\mathrm{A}^m|=q^m$. Let $\mathrm{A}_w(n)$ denote the number of words of length n that do not contain the factor $w\in\mathrm{A}^*$. Let $\mu(n,m)$ be the maximal value of $\mathrm{A}_w(n)$ for all w of length m; formally

$$\mu(n,m)=\max\{\mathrm{A}_w(n)\mid w\in\mathrm{A}^m\}.$$

Let $\hat{\mathrm{D}}(n)$ denote the set of all closed words of length n and let $\hat{\mathrm{D}}(n,m)$ denote the set of all closed words of length n having a maximal border of length m. Let $\mathrm{D}(n)=|\hat{\mathrm{D}}(n)|$ and $\mathrm{D}(n,m)=|\hat{\mathrm{D}}(n,m)|$.

Obviously $\hat{\mathrm{D}}(n)=\bigcup_{m=1}^{n-1}\hat{\mathrm{D}}(n,m)$ and $\hat{\mathrm{D}}(n,m)\cap\hat{\mathrm{D}}(n,\bar{m})=\emptyset$, where $m\neq\bar{m}$. We show that if $2m>n$ then $\mathrm{D}(n,m)\leq q^{\lceil\frac{n}{2}\rceil}$ and if $2m\leq n$ then $\mathrm{D}(n,m)\leq q^m\mu(n-2m,m)$; see Lemma 2.5. It follows that

$$\mathrm{D}(n)=\sum_{m=1}^{n-1}\mathrm{D}(n,m)\leq\sum_{m=1}^{\lfloor\frac{n}{2}\rfloor}q^m\mu(n-2m,m)+\sum_{m=\lfloor\frac{n}{2}\rfloor+1}^{n-1}q^{\lceil\frac{n}{2}\rceil}. \tag{1}$$

Let $\mathbb{N}$ denote the set of positive integers. Let $\omega(n)=\frac{1}{\ln q}(\ln n-\ln\ln n)$. Let Π denote the set of all functions $\pi(n):\mathbb{N}\to\mathbb{N}$ such that $\pi(n)\in\Pi$ if and only if $1\leq\pi(n)\leq\max\{1,\omega(n)\}$ and $\pi(n)\leq\pi(n+1)$ for all $n\in\mathbb{N}$. We apply the function max, because $\omega(n)<1$ for some small n.

The key observation in our article is that the number of words of length n that do not contain some "short" factor of length $\pi(n)\in\Pi$ has the same growth rate as the number of words of length $n-\lfloor\frac{\ln n}{\ln q}\rfloor$. Formally said, for each $\pi(n)\in\Pi$ there is a positive real constant c such that $\mu(n,\pi(n))\leq cq^{n-\frac{\ln n}{\ln q}}$; see Theorem 2.3. This observation allows us to show that there are real positive constants c_1,c_2 such that

$$\sum_{m=1}^{\lfloor\frac{n}{2}\rfloor}q^m\mu(n-2m,m)\leq c_1\ln n\sum_{m=\lfloor c_2\ln n\rfloor}^{\lfloor\frac{n}{2}\rfloor}q^m\mu(n-2m,m). \tag{2}$$

In consequence we may count only closed words having a maximal border longer than $c_2\ln n$ in order to find an upper bound for $\mathrm{D}(n)$.

Applying that $\mu(n-2m,m)\leq q^{n-2m}$ for $n\geq 2m$, we derive from (1) and (2) our result for the number of closed words.

2. Upper bound for the number of closed words

We present an upper bound for the number of words of length n that avoid some factor of length m; it means an upper bound for $\mu(n,m)$.

Lemma 2.1. *If $n,m\in\mathbb{N}$ then*

$$\mu(n,m)\leq q^n\left(1-\frac{1}{q^m}\right)^{\lfloor\frac{n}{m}\rfloor}.$$

Proof. Given $w\in\mathrm{A}^m$, let $U_{n,w}$ be a set of words $u=u_1u_2\dots u_{k-1}u_k\in\mathrm{A}^*$, where $|u|=n$, $|u_i|=m$, $w\neq u_i$ for all $1\leq i<k$, and $|u_k|=n \bmod m$. It follows that $|u_k|<m=|w|$ and thus $u_k\neq w$. Obviously

$$|U_{n,w}|=(q^m-1)^{\lfloor\frac{n}{m}\rfloor}q^{n \bmod m}=q^n\left(1-\frac{1}{q^m}\right)^{\lfloor\frac{n}{m}\rfloor}.$$

Note that $|\mathrm{A}^m\setminus\{w\}|=q^m-1$. It is clear that the set of words of length n not containing the factor w is a subset of $U_{n,w}$. The lemma follows. □

For the proof of Theorem 2.3 we need the following limit.

Proposition 2.2. *We have that*

$$\lim_{n\to\infty}n\left(1-\frac{\ln n}{n}\right)^n=\mathrm{e}.$$

Proof. Let

$$y=\lim_{n\to\infty}n\left(1-\frac{\ln n}{n}\right)^n. \tag{3}$$

From (3) we have that

$$\ln y = \lim_{n\to\infty} \ln\left[n\left(1-\frac{\ln n}{n}\right)^n\right] = \lim_{n\to\infty}\left[\ln n + n\ln\left(1-\frac{\ln n}{n}\right)\right]. \tag{4}$$

Let us consider the second term on the right side of (4):

$$\lim_{n\to\infty} n\ln\left(1-\frac{\ln n}{n}\right) = \lim_{n\to\infty}\frac{\ln\left(1-\frac{\ln n}{n}\right)'}{(\frac{1}{n})'} = \lim_{n\to\infty}\frac{\frac{(-1)(\frac{1-\ln n}{n^2})}{\left(1-\frac{\ln n}{n}\right)}}{-\frac{1}{n^2}} = \lim_{n\to\infty}\frac{n(1-\ln n)}{n-\ln n}. \tag{5}$$

Since $\lim_{n\to\infty}\frac{n}{n-\ln n} = 1$, it follows from (4) and (5) that

$$\ln y = \lim_{n\to\infty}\left[\ln n + \frac{n(1-\ln n)}{n-\ln n}\right] = \lim_{n\to\infty}[\ln n + 1 - \ln n] = 1.$$

It follows that $y = \mathrm{e}$. This completes the proof. □

Let $\mathbb{R}^+$ denote the set of positive real numbers.

Let $\beta = \frac{1}{\ln q} \in \mathbb{R}^+$. The following theorem states that the number of words of length n avoiding some given "short" factor (of length shorter than $\pi(n) \in \Pi$) has the same growth rate as the number of all words of length $n - \beta\ln n$.

Theorem 2.3. *If $\pi(n) \in \Pi$ then there is a constant $c \in \mathbb{R}^+$ such that for all $n \in \mathbb{N}$ we have that*

$$\frac{\mu(n,\pi(n))}{q^{n-\beta\ln n}} \le c.$$

Proof. From Lemma 2.1 we have that

$$\frac{\mu(n,\pi(n))}{q^{n-\beta\ln n}} \le \frac{q^n\left(1-\frac{1}{q^{\pi(n)}}\right)^{\lfloor\frac{n}{\pi(n)}\rfloor}}{q^{n-\beta\ln n}} = n\left(1-\frac{1}{q^{\pi(n)}}\right)^{\lfloor\frac{n}{\pi(n)}\rfloor}. \tag{6}$$

Realize that $q^{\beta\ln n} = n$.

Obviously there is $n_0 \in \mathbb{N}$ such that $q^{\pi(n)} \le \frac{n}{\ln n}$ for all $n > n_0$; recall that $\pi(n) \le \omega(n) = \frac{1}{\ln q}(\ln n - \ln\ln n)$ as n tends to infinity. Consequently for all $n > n_0$ we have that

$$n\left(1-\frac{1}{q^{\pi(n)}}\right)^n \le n\left(1-\frac{\ln n}{n}\right)^n. \tag{7}$$

Proposition 2.2 and (7) imply that

$$\lim_{n\to\infty} n\left(1-\frac{1}{q^{\pi(n)}}\right)^n \le \mathrm{e}. \tag{8}$$

Clearly $\lim_{n\to\infty}(f(n))^{\frac{1}{\pi(n)}} \le \mathrm{e}$ for each function $f(n)$ such that $f(n) \ge 0$ and $\lim_{n\to\infty} f(n) \le \mathrm{e}$; recall that $\pi(n) \ge 1$. Then the theorem follows from (6) and (8). This completes the proof. □

Let $h(n) = \lfloor\beta\ln n\rfloor$. We present Theorem 2.3 in a slightly different manner that will be more useful for us in the following.

Corollary 2.4. *If $\bar{\pi}(n) \in \Pi$ then there is a constant $c \in \mathbb{R}^+$ such that for all $n \in \mathbb{N}$ we have that*

$$\frac{\mu(n-2\bar{\pi}(n),\bar{\pi}(n))}{q^{n-h(n)}} \le c.$$

Proof. Let $\pi(n) \in \Pi$ be a function such that $\bar{\pi}(n) \le \pi(n)$. It is easy to verify that $\mu(n-2\bar{\pi}(n),\bar{\pi}(n)) \le \mu(n,\pi(n))$, since the number of words of length n avoiding some factor of length $\pi(n)$ is bigger or equal to the number of words of length $n-2\bar{\pi}(n)$ avoiding some factor of length $\bar{\pi}(n) \le \pi(n)$.

Obviously $h(n) = \lfloor\frac{\ln n}{\ln q}\rfloor \le \frac{\ln n}{\ln q} = \beta\ln n$. In consequence we have that $q^{n-h(n)} \ge q^{n-\beta\ln n}$.

The corollary follows from Theorem 2.3. This completes the proof. □

We show an upper bound for $\mathrm{D}(n,m)$ for the cases where $2m > n$ and $2m \leq n$.

Lemma 2.5. *Suppose $n, m \in \mathbb{N}$.*

- *If $2m > n$ then $\mathrm{D}(n,m) \leq q^{\lceil \frac{n}{2} \rceil}$.*
- *If $2m \leq n$ then $\mathrm{D}(n,m) \leq q^m \mu(n-2m, m)$.*

Proof. If $2m > n$, $w \in \mathrm{A}^*$, and $|w| = m$ then there is obviously at most one word u with $|u| = n$ having a prefix and a suffix w; the prefix w and the suffix w would overlap with each other. If such u exists then the first half of u uniquely determines the second half of u. It follows that $\mathrm{D}(n,m) \leq q^{\lceil \frac{n}{2} \rceil}$.

Let $\mathrm{F}(w)$ denote the set of all factors of $w \in \mathrm{A}^*$. If $n \geq 2m$ then let

$$Z(n,m) = \{wuw \mid u \in \mathrm{A}^{n-2m} \text{ and } w \in \mathrm{A}^m \text{ and } w \notin \mathrm{F}(u)\}.$$

If $n \geq 2m$ then $\hat{\mathrm{D}}(n,m) \subseteq Z(n,m)$. It is easy to see that

$$|Z(n,m)| \leq |\mathrm{A}^m| \mu(n-2m, m).$$

This completes the proof. $\square$

Let $\kappa > 1$ be a real constant and $\bar{h}(n) = \max\{1, \lfloor \frac{1}{\kappa}\omega(n) \rfloor\}$. Again we use the function max to guarantee that $\bar{h}(n) \geq 1$ for small n.

Remark 2.6. The function $\bar{h}(n)$ defines the maximal length of a "short" border of a closed word. In the proof of Theorem 2.9 the closed words from $\hat{\mathrm{D}}(n,m)$ will be enumerated differently for $m < \bar{h}(n)$ and for $m \geq \bar{h}(n)$.

The next auxiliary lemma shows an upper bound for $q^{-h(n)+\bar{h}(n)}$, that we will use in the proof of Proposition 2.8.

Lemma 2.7. *There is a constant $c_1 \in \mathbb{R}^+$ such that for all $n \in \mathbb{N}$ we have that*

$$q^{-h(n)+\bar{h}(n)} \leq c_1 q^{\frac{1}{\ln q}\left(\frac{1}{\kappa}-1\right)\ln n}.$$

Proof. Let

$$y = \lim_{n\to\infty} (-h(n) + \bar{h}(n) - \frac{1}{\ln q}\left(\frac{1}{\kappa} - 1\right)\ln n).$$

We have that

$$\begin{aligned} y &= \lim_{n\to\infty}\left(-\left\lfloor \frac{1}{\ln q}\ln n \right\rfloor + \left\lfloor \frac{1}{\kappa \ln q}(\ln n - \ln\ln n)\right\rfloor - \frac{1}{\ln q}\left(\frac{1}{\kappa}-1\right)\ln n\right) \\ &= \lim_{n\to\infty}\left(\frac{\ln n}{\ln q}\left(-1+\frac{1}{\kappa}\right) - \frac{1}{\ln q}\left(\frac{1}{\kappa}-1\right)\ln n - \frac{1}{\kappa \ln q}\ln\ln n\right) \\ &= -\infty. \end{aligned} \tag{9}$$

This implies that

$$\lim_{n\to\infty} \frac{q^{-h(n)+\bar{h}(n)}}{q^{\frac{1}{\ln q}\left(\frac{1}{\kappa}-1\right)\ln n}} = q^y = q^{-\infty} = 0.$$

The lemma follows. $\square$

The next proposition shows an upper bound for the number of closed words of length n having a maximal border of length $\leq \lceil \frac{n}{2} \rceil$.

Proposition 2.8. *There is a constant $c \in \mathbb{R}^+$ such that*

$$\sum_{m=1}^{\lceil \frac{n}{2} \rceil} q^m \mu(n-2m, m) \leq c \ln n \frac{q^n}{\sqrt{n}}, \text{ where } n > 1.$$

Proof. Since $\mu(n-2m,m)\le q^{n-2m}$ we have that

$$\sum_{m=1}^{\lceil \frac{n}{2}\rceil} q^m\mu(n-2m,m)\le \sum_{m=1}^{\bar{h}(n)-1} q^m\mu(n-2m,m)+\sum_{m=\bar{h}(n)}^{\lceil \frac{n}{2}\rceil} q^m q^{n-2m}. \tag{10}$$

Corollary 2.4 implies that $\mu(n-2m,m)\le cq^{n-h(n)}$ for some constant $c\in\mathbb{R}^+$. It follows that

$$\begin{aligned}\sum_{m=1}^{\bar{h}(n)-1} q^m\mu(n-2m,m)&\le \sum_{m=1}^{\bar{h}(n)} q^m cq^{n-h(n)}\\ &\le \bar{h}(n)q^{\bar{h}(n)}cq^{n-h(n)}.\end{aligned} \tag{11}$$

Lemma 2.7 and (11) imply that

$$\sum_{m=1}^{\bar{h}(n)-1} q^m\mu(n-2m,m)\le c_1\bar{h}(n)cq^{n-\frac{\ln n}{\ln q}(1-\frac{1}{\kappa})}, \tag{12}$$

where c_1 is some real positive constant.

It is easy to verify that

$$q^{-\bar{h}(n)}\le q^{-\frac{1}{\kappa\ln q}(\ln n-\ln\ln n)+1}=q(\ln n)^{\frac{1}{\kappa}}q^{-\frac{1}{\kappa\ln q}\ln n}. \tag{13}$$

Thus using (13)

$$\sum_{m=\bar{h}(n)}^{\lceil \frac{n}{2}\rceil} q^m q^{n-2m}\le q^n\sum_{m=\bar{h}(n)}^{\lceil \frac{n}{2}\rceil} q^{-m}\le \frac{q^{n-\bar{h}(n)}}{1-q^{-1}}\le \frac{q(\ln n)^{\frac{1}{\kappa}}q^{n-\frac{1}{\kappa\ln q}\ln n}}{1-q^{-1}}. \tag{14}$$

Obviously $\bar{h}(n)\le \frac{\ln n}{\kappa\ln q}$. Hence taking $\kappa=2$, we get from (10), (12), and (14) that

$$\begin{aligned}\sum_{m=1}^{\lceil \frac{n}{2}\rceil} q^m\mu(n-2m,m)&\le c_1\bar{h}(n)cq^{n-\frac{1}{2\ln q}\ln n}+\frac{q(\ln n)^{\frac{1}{2}}q^{n-\frac{1}{2\ln q}\ln n}}{1-q^{-1}}\\ &\le q^{n-\frac{1}{2\ln q}\ln n}\left(c_1c\frac{\ln n}{2\ln q}+\frac{q(\ln n)^{\frac{1}{2}}}{1-q^{-1}}\right)\\ &\le q^{n-\frac{1}{2\ln q}\ln n}(c_2\ln n+c_3(\ln n)^{\frac{1}{2}}),\end{aligned} \tag{15}$$

for some constants $c_2,c_3\in\mathbb{R}^+$. Since $\sqrt{n}=q^{\frac{1}{2\ln q}\ln n}$ the proposition follows from (15). □

We show an upper bound for $\mathrm{D}(n)$.

Theorem 2.9. *There is a constant $c\in\mathbb{R}^+$ such that*

$$\mathrm{D}(n)\le c\ln n\frac{q^n}{\sqrt{n}},\ \text{where } n>1.$$

Proof. We have that

$$\mathrm{D}(n)=\sum_{m=1}^{n-1}\mathrm{D}(n,m)=\sum_{m=1}^{\lceil \frac{n}{2}\rceil}\mathrm{D}(n,m)+\sum_{m=\lceil \frac{n}{2}\rceil+1}^{n-1}\mathrm{D}(n,m). \tag{16}$$

From Lemma 2.5 and (16) we get that

$$\mathrm{D}(n)\le \sum_{m=1}^{\lceil \frac{n}{2}\rceil} q^m\mu(n-2m,m)+\sum_{m=\lceil \frac{n}{2}\rceil+1}^{n-1} q^{\lceil \frac{n}{2}\rceil}. \tag{17}$$

Realize that

$$\sum_{m=\lceil \frac{n}{2} \rceil+1}^{n-1} q^{\lceil \frac{n}{2} \rceil} \leq \frac{n}{2} q^{\lceil \frac{n}{2} \rceil}$$

and

$$\lim_{n\to\infty} \frac{nq^{\frac{n}{2}}}{\ln n \frac{q^n}{\sqrt{n}}} = 0.$$

Then it follows that from (17), and Proposition 2.8 that there are constants $c_2, c_3 \in \mathbb{R}^+$ such that

$$c_2 \sum_{m=1}^{\lceil \frac{n}{2} \rceil} q^m \mu(n-2m, m) \geq \sum_{m=\lceil \frac{n}{2} \rceil+1}^{n-1} q^{\lceil \frac{n}{2} \rceil} \text{ and}$$

$$\mathrm{D}(n) \leq c_3 \sum_{m=1}^{\lceil \frac{n}{2} \rceil} q^m \mu(n-2m, m). \tag{18}$$

The theorem follows from (18), and Proposition 2.8. □

Remark 2.10. Note that some of the constants c, c_1, c_2, c_3, that we used in our results and in particular in Theorem 2.9, depend on q.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author acknowledges support by the Czech Science Foundation grant GAČR 13-03538S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/205/OHK4/3T/14.

References

[1] M. Forsyth, A. Jayakumar, J. Peltomäki, J. Shallit, Remarks on privileged words, Int. J. Found. Comput. Sci. 27 (04) (2016) 431–442, https://doi.org/10.1142/S0129054116500088.
[2] J. Kellendonk, D. Lenz, J. Savinien, A characterization of subshifts with bounded powers, Discrete Math. 313 (24) (2013) 2881–2894, https://doi.org/10.1016/j.disc.2013.08.026.
[3] J. Nicholson, N. Rampersad, Improved estimates for the number of privileged words, J. Integer Seq. 21 (2018).
[4] J. Peltomäki, Privileged words and Sturmian words, TUCS Dissertations No. 214, Turku Centre for Computer Science, August 2016, available at http://urn.fi/URN:ISBN:978-952-12-3422-4.
[5] J. Peltomäki, Introducing privileged words: privileged complexity of Sturmian words, Theor. Comput. Sci. 500 (2013) 57–67, https://doi.org/10.1016/j.tcs.2013.05.028.
[6] L. Schaeffer, J. Shallit, Closed, palindromic, rich, privileged, trapezoidal, and balanced words in automatic sequences, Electron. J. Comb. 23 (2016) P1.25.

Article [[Ru07]]: Transition Property for α-power-free Languages with $\alpha \geq 2$ and $k \geq 3$ Letters

Transition Property for α-Power Free Languages with $\alpha \geq 2$ and $k \geq 3$ Letters

Josef Rukavicka(✉)

Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic
josef.rukavicka@seznam.cz

Abstract. In 1985, Restivo and Salemi presented a list of five problems concerning power free languages. Problem 4 states: Given α-power-free words u and v, decide whether there is a transition from u to v. Problem 5 states: Given α-power-free words u and v, find a transition word w, if it exists.

Let Σ_k denote an alphabet with k letters. Let $L_{k,\alpha}$ denote the α-power free language over the alphabet Σ_k, where α is a rational number or a rational "number with +". If α is a "number with +" then suppose $k \geq 3$ and $\alpha \geq 2$. If α is "only" a number then suppose $k = 3$ and $\alpha > 2$ or $k > 3$ and $\alpha \geq 2$. We show that: If $u \in L_{k,\alpha}$ is a right extendable word in $L_{k,\alpha}$ and $v \in L_{k,\alpha}$ is a left extendable word in $L_{k,\alpha}$ then there is a (transition) word w such that $uwv \in L_{k,\alpha}$. We also show a construction of the word w.

Keywords: Power free languages · Transition property · Dejean's conjecture

1 Introduction

The power free words are one of the major themes in the area of combinatorics on words. An *α-power* of a word r is the word $r^\alpha = rr\dots rt$ such that $\frac{|r^\alpha|}{|r|} = \alpha$ and t is a prefix of r, where $\alpha \geq 1$ is a rational number. For example $(1234)^3 = 123412341234$ and $(1234)^{\frac{7}{4}} = 1234123$. We say that a finite or infinite word w is *α-power free* if w has no factors that are β-powers for $\beta \geq \alpha$ and we say that a finite or infinite word w is α^+-power free if w has no factors that are β-powers for $\beta > \alpha$, where $\alpha, \beta \geq 1$ are rational numbers. In the following, when we write "α-power free" then α denotes a number or a "number with +". The power free words, also called repetitions free words, include well known square free (2-power free), overlap free (2^+-power free), and cube free words (3-power free). Two surveys on the topic of power free words can be found in [8] and [13].

One of the questions being researched is the construction of infinite power free words. We define the *repetition threshold* $\mathrm{RT}(k)$ to be the infimum of all rational numbers α such that there exists an infinite α-power-free word over an alphabet with k letters. Dejean's conjecture states that $\mathrm{RT}(2) = 2$, $\mathrm{RT}(3) = \frac{7}{4}$,

N. Jonoska and D. Savchuk (Eds.): DLT 2020, LNCS 12086, pp. 294–303, 2020.
https://doi.org/10.1007/978-3-030-48516-0_22

$\mathrm{RT}(4) = \frac{7}{5}$, and $\mathrm{RT}(k) = \frac{k}{k-1}$ for each $k > 4$ [3]. Dejean's conjecture has been proved with the aid of several articles [1–3,5,6,9].

It is easy to see that α-power free words form a factorial language [13]; it means that all factors of a α-power free word are also α-power free words. Then Dejean's conjecture implies that there are infinitely many finite α-power free words over Σ_k, where $\alpha > \mathrm{RT}(k)$.

In [10], Restivo and Salemi presented a list of five problems that deal with the question of extendability of power free words. In the current paper we investigate Problem 4 and Problem 5:

– Problem 4: Given α-power-free words u and v, decide whether there is a transition word w, such that uwu is α-power free.
– Problem 5: Given α-power-free words u and v, find a transition word w, if it exists.

A recent survey on the progress of solving all the five problems can be found in [7]; in particular, the problems 4 and 5 are solved for some overlap free (2^+-power free) binary words. In addition, in [7] the authors prove that: For every pair (u, v) of cube free words (3-power free) over an alphabet with k letters, if u can be infinitely extended to the right and v can be infinitely extended to the left respecting the cube-freeness property, then there exists a "transition" word w over the same alphabet such that uwv is cube free.

In 2009, a conjecture related to Problems 4 and Problem 5 of Restivo and Salemi appeared in [12]:

Conjecture 1. [12, Conjecture 1] Let L be a power-free language and let $e(L) \subseteq L$ be the set of words of L that can be extended to a bi-infinite word respecting the given power-freeness. If $u, v \in e(L)$ then $uwv \in e(L)$ for some word w.

In 2018, Conjecture 1 was presented also in [11] in a slightly different form.

Let $\mathbb{N}$ denote the set of natural numbers and let $\mathbb{Q}$ denote the set of rational numbers.

Definition 1. *Let*

$$\begin{aligned} \Upsilon = &\{(k, \alpha) \mid k \in \mathbb{N} \textit{ and } \alpha \in \mathbb{Q} \textit{ and } k = 3 \textit{ and } \alpha > 2\} \\ &\cup \{(k, \alpha) \mid k \in \mathbb{N} \textit{ and } \alpha \in \mathbb{Q} \textit{ and } k > 3 \textit{ and } \alpha \geq 2\} \\ &\cup \{(k, \alpha^+) \mid k \in \mathbb{N} \textit{ and } \alpha \in \mathbb{Q} \textit{ and } k \geq 3 \textit{ and } \alpha \geq 2\}. \end{aligned}$$

Remark 1. The definition of Υ says that: If $(k, \alpha) \in \Upsilon$ and α is a "number with +" then $k \geq 3$ and $\alpha \geq 2$. If $(k, \alpha) \in \Upsilon$ and α is "just" a number then $k = 3$ and $\alpha > 2$ or $k > 3$ and $\alpha \geq 2$.

Let L be a language. A finite word $w \in$ L is called *left extendable* (resp., *right extendable*) in L if for every $n \in \mathbb{N}$ there is a word $u \in$ L with $|u| = n$ such that $uw \in$ L (resp., $wu \in$ L).

In the current article we improve the results addressing Problems 4 and Problem 5 of Restivo and Salemi from [7] as follows. Let Σ_k denote an alphabet with k letters. Let $\mathrm{L}_{k,\alpha}$ denote the α-power free language over the alphabet Σ_k. We show that if $(k,\alpha) \in \Upsilon$, $u \in \mathrm{L}_{k,\alpha}$ is a right extendable word in $\mathrm{L}_{k,\alpha}$, and $v \in \mathrm{L}_{k,\alpha}$ is a left extendable word in $\mathrm{L}_{k,\alpha}$ then there is a word w such that $uwv \in \mathrm{L}_{k,\alpha}$. We also show a construction of the word w.

We sketch briefly our construction of a "transition" word. Let u be a right extendable α-power free word and let v be a left extendable α-power free word over Σ_k with $k > 2$ letters. Let $\bar{u}$ be a right infinite α-power free word having u as a prefix and let $\bar{v}$ be a left infinite α-power free word having v as a suffix. Let x be a letter that is recurrent in both $\bar{u}$ and $\bar{v}$. We show that we may suppose that $\bar{u}$ and $\bar{v}$ have a common recurrent letter. Let t be a right infinite α-power free word over $\Sigma_k \setminus \{x\}$. Let $\bar{t}$ be a left infinite α-power free word such that the set of factors of $\bar{t}$ is a subset of the set of recurrent factors of t. We show that such $\bar{t}$ exists. We identify a prefix $\tilde{u}xg$ of $\bar{u}$ such that g is a prefix of t and $\tilde{u}xt$ is a right infinite α-power free word. Analogously we identify a suffix $\bar{g}x\tilde{v}$ of $\bar{v}$ such that $\bar{g}$ is a suffix of $\bar{t}$ and $\bar{t}x\tilde{v}$ is a left infinite α-power free word. Moreover our construction guarantees that u is a prefix of $\tilde{u}xt$ and v is a suffix of $\bar{t}x\tilde{v}$. Then we find a prefix hp of t such that $px\tilde{v}$ is a suffix of $\bar{t}x\tilde{v}$ and such that both h and p are "sufficiently long". Then we show that $\tilde{u}xhpx\tilde{v}$ is an α-power free word having u as a prefix and v as a suffix.

The very basic idea of our proof is that if u, v are α-power free words and x is a letter such that x is not a factor of both u and v, then clearly uxv is α-power free on condition that $\alpha \geq 2$. Just note that there cannot be a factor in uxv which is an α-power and contains x, because x has only one occurrence in uxv. Our constructed words $\tilde{u}xt$, $\bar{t}x\tilde{v}$, and $\tilde{u}xhpx\tilde{v}$ have "long" factors which does not contain a letter x. This will allow us to apply a similar approach to show that the constructed words do not contain square factor rr such that r contains the letter x.

Another key observation is that if $k \geq 3$ and $\alpha > \mathrm{RT}(k-1)$ then there is an infinite α-power free word $\bar{w}$ over $\Sigma_k \setminus \{x\}$, where $x \in \Sigma_k$. This is an implication of Dejean's conjecture. Less formally said, if u, v are α-power free words over an alphabet with k letters, then we construct a "transition" word w over an alphabet with $k-1$ letters such that uwv is α-power free.

Dejean's conjecture imposes also the limit to possible improvement of our construction. The construction cannot be used for $\mathrm{RT}(k) \leq \alpha < \mathrm{RT}(k-1)$, where $k \geq 3$, because every infinite (or "sufficiently long") word w over an alphabet with $k-1$ letters contains a factor which is an α-power. Also for $k = 2$ and $\alpha \geq 1$ our technique fails. On the other hand, based on our research, it seems that our technique, with some adjustments, could be applied also for $\mathrm{RT}(k-1) \leq \alpha \leq 2$ and $k \geq 3$. Moreover it seems to be possible to generalize our technique to bi-infinite words and consequently to prove Conjecture 1 for $k \geq 3$ and $\alpha \geq \mathrm{RT}(k-1)$.

2 Preliminaries

Recall that Σ_k denotes an alphabet with k letters. Let ϵ denote the empty word. Let Σ_k^* denote the set of all finite words over Σ_k including the empty word ϵ, let

$\Sigma_k^{\mathbb{N},R}$ denote the set of all right infinite words over Σ_k, and let $\Sigma_k^{\mathbb{N},L}$ denote the set of all left infinite words over Σ_k. Let $\Sigma_k^{\mathbb{N}} = \Sigma_k^{\mathbb{N},L} \cup \Sigma_k^{\mathbb{N},R}$. We call $w \in \Sigma_k^{\mathbb{N}}$ an infinite word.

Let $\mathrm{occur}(w,t)$ denote the number of occurrences of the nonempty factor $t \in \Sigma_k^* \setminus \{\epsilon\}$ in the word $w \in \Sigma_k^* \cup \Sigma_k^{\mathbb{N}}$. If $w \in \Sigma_k^{\mathbb{N}}$ and $\mathrm{occur}(w,t) = \infty$, then we call t a *recurrent* factor in w.

Let $\mathrm{F}(w)$ denote the set of all finite factors of a finite or infinite word $w \in \Sigma_k^* \cup \Sigma_k^{\mathbb{N}}$. The set $\mathrm{F}(w)$ contains the empty word and if w is finite then also $w \in \mathrm{F}(w)$. Let $\mathrm{F}_r(w) \subseteq \mathrm{F}(w)$ denote the set of all recurrent nonempty factors of $w \in \Sigma_k^{\mathbb{N}}$.

Let $\mathrm{Prf}(w) \subseteq \mathrm{F}(w)$ denote the set of all prefixes of $w \in \Sigma_k^* \cup \Sigma_k^{\mathbb{N},R}$ and let $\mathrm{Suf}(w) \subseteq \mathrm{F}(w)$ denote the set of all suffixes of $w \in \Sigma_k^* \cup \Sigma_k^{\mathbb{N},L}$. We define that $\epsilon \in \mathrm{Prf}(w) \cap \mathrm{Suf}(w)$ and if w is finite then also $w \in \mathrm{Prf}(w) \cap \mathrm{Suf}(w)$.

We have that $\mathrm{L}_{k,\alpha} \subseteq \Sigma_k^*$. Let $\mathrm{L}_{k,\alpha}^{\mathbb{N}} \subseteq \Sigma_k^{\mathbb{N}}$ denote the set of all infinite α-power free words over Σ_k. Obviously $\mathrm{L}_{k,\alpha}^{\mathbb{N}} = \{w \in \Sigma_k^{\mathbb{N}} \mid \mathrm{F}(w) \subseteq \mathrm{L}_{k,\alpha}\}$. In addition we define $\mathrm{L}_{k,\alpha}^{\mathbb{N},R} = \mathrm{L}_{k,\alpha}^{\mathbb{N}} \cap \Sigma_k^{\mathbb{N},R}$ and $\mathrm{L}_{k,\alpha}^{\mathbb{N},L} = \mathrm{L}_{k,\alpha}^{\mathbb{N}} \cap \Sigma_k^{\mathbb{N},L}$; it means the sets of right infinite and left infinite α-power free words.

3 Power Free Languages

Let $(k,\alpha) \in \Upsilon$ and let u, v be α-power free words. The first lemma says that uv is α-power free if there are no word r and no nonempty prefix $\bar{v}$ of v such that rr is a suffix of $u\bar{v}$ and rr is longer than $\bar{v}$.

Lemma 1. *Suppose $(k,\alpha) \in \Upsilon$, $u \in \mathrm{L}_{k,\alpha}$, and $v \in \mathrm{L}_{k,\alpha} \cup \mathrm{L}_{k,\alpha}^{\mathbb{N},R}$. Let*

$$\Pi = \{(r,\bar{v}) \mid r \in \Sigma_k^* \setminus \{\epsilon\} \text{ and } \bar{v} \in \mathrm{Prf}(v) \setminus \{\epsilon\} \text{ and } rr \in \mathrm{Suf}(u\bar{v}) \text{ and } |rr| > |\bar{v}|\}.$$

If $\Pi = \emptyset$ then $uv \in \mathrm{L}_{k,\alpha} \cup \mathrm{L}_{k,\alpha}^{\mathbb{N},R}$.

Proof. Suppose that uv is not α-power free. Since u is α-power free, then there are $t \in \Sigma_k^*$ and $x \in \Sigma_k$ such that $tx \in \mathrm{Prf}(v)$, $ut \in \mathrm{L}_{k,\alpha}$ and $utx \notin \mathrm{L}_{k,\alpha}$. It means that there is $r \in \mathrm{Suf}(utx)$ such that $r^\beta \in \mathrm{Suf}(utx)$ for some $\beta \geq \alpha$ or $\beta > \alpha$ if α is a "number with +"; recall Definition 1 of Υ. Because $\alpha \geq 2$, this implies that $rr \in \mathrm{Suf}(r^\beta)$. If follows that $(tx, r) \in \Pi$. We proved that $uv \notin \mathrm{L}_{k,\alpha} \cup \mathrm{L}_{k,\alpha}^{\mathbb{N},R}$ implies that $\Pi \neq \emptyset$. The lemma follows. $\square$

The following technical set $\Gamma(k,\alpha)$ of 5-tuples (w_1, w_2, x, g, t) will simplify our propositions.

Definition 2. *Given $(k,\alpha) \in \Upsilon$, we define that $(w_1, w_2, x, g, t) \in \Gamma(k,\alpha)$ if*

1. *$w_1, w_2, g \in \Sigma_k^*$,*
2. *$x \in \Sigma_k$,*
3. *$w_1 w_2 x g \in \mathrm{L}_{k,\alpha}$,*

4. $t \in \mathrm{L}_{k,\alpha}^{\mathbb{N},R}$*,*
5. $\mathrm{occur}(t, x) = 0$*,*
6. $g \in \mathrm{Prf}(t)$*,*
7. $\mathrm{occur}(w_2xgy, xgy) = 1$*, where* $y \in \Sigma_k$ *is such that* $gy \in \mathrm{Prf}(t)$*, and*
8. $\mathrm{occur}(w_2, x) \geq \mathrm{occur}(w_1, x)$*.*

Remark 2. Less formally said, the 5-tuple (w_1, w_2, x, g, t) is in $\Gamma(k, \alpha)$ if w_1w_2xg is α-power free word over Σ_k, t is a right infinite α-power free word over Σ_k, t has no occurrence of x (thus t is a word over $\Sigma_k \setminus \{x\}$), g is a prefix of t, xgy has only one occurrence in w_2xgy, where y is a letter such that gy is a prefix of t, and the number of occurrences of x in w_2 is bigger than the number of occurrences of x in w_1, where w_1, w_2, g are finite words and x is a letter.

The next proposition shows that if (w_1, w_2, x, g, t) is from the set $\Gamma(k, \alpha)$ then w_1w_2xt is a right infinite α-power free word, where (k, α) is from the set Υ.

Proposition 1. *If* $(k, \alpha) \in \Upsilon$ *and* $(w_1, w_2, x, g, t) \in \Gamma(k, \alpha)$ *then* $w_1w_2xt \in \mathrm{L}_{k,\alpha}^{\mathbb{N},R}$.

Proof. Lemma 1 implies that it suffices to show that there are no $u \in \mathrm{Prf}(t)$ with $|u| > |g|$ and no $r \in \Sigma_k^* \setminus \{\epsilon\}$ such that $rr \in \mathrm{Suf}(w_1w_2xu)$ and $|rr| > |u|$. Recall that w_1w_2xg is an α-power free word, hence we consider $|u| > |g|$. To get a contradiction, suppose that such r, u exist. We distinguish the following distinct cases.

- If $|r| \leq |u|$ then: Since $u \in \mathrm{Prf}(t) \subseteq \mathrm{L}_{k,\alpha}$ it follows that $xu \in \mathrm{Suf}(r^2)$ and hence $x \in \mathrm{F}(r^2)$. It is clear that $\mathrm{occur}(r^2, x) \geq 1$ if and only if $\mathrm{occur}(r, x) \geq 1$. Since $x \notin \mathrm{F}(u)$ and thus $x \notin \mathrm{F}(r)$, this is a contradiction.
- If $|r| > |u|$ and $rr \in \mathrm{Suf}(w_2xu)$ then: Let $y \in \Sigma_k$ be such that $gy \in \mathrm{Prf}(t)$. Since $|u| > |g|$ we have that $gy \in \mathrm{Prf}(u)$ and $xgy \in \mathrm{Prf}(xu)$. Since $|r| > |u|$ we have that $xgy \in \mathrm{F}(r)$. In consequence $\mathrm{occur}(rr, xgy) \geq 2$. But Property 7 of Definition 2 states that $\mathrm{occur}(w_2xgy, xgy) = 1$. Since $rr \in \mathrm{Suf}(w_2xu)$, this is a contradiction.
- If $|r| > |u|$ and $rr \notin \mathrm{Suf}(w_2xu)$ and $r \in \mathrm{Suf}(w_2xu)$ then:
  Let $w_{11}, w_{12}, w_{13}, w_{21}, w_{22} \in \Sigma_k^*$ be such that $w_1 = w_{11}w_{12}w_{13}$, $w_2 = w_{21}w_{22}$, $w_{12}w_{13}w_{21} = r$, $w_{12}w_{13}w_2xu = rr$, and $w_{13}w_{21} = xu$; see Figure below.

<table>
<tr><td></td><td></td><td colspan="2">xu</td><td></td><td></td><td></td></tr>
<tr><td>w_{11}</td><td>w_{12}</td><td>w_{13}</td><td>w_{21}</td><td>w_{22}</td><td>x</td><td>u</td></tr>
<tr><td></td><td colspan="3">r</td><td colspan="3">r</td></tr>
</table>

  It follows that $w_{22}xu = r$ and $w_{22} = w_{12}$. It is easy to see that $w_{13}w_{21} = xu$. From $\mathrm{occur}(u, x) = 0$ we have that $\mathrm{occur}(w_2, x) = \mathrm{occur}(w_{22}, x)$ and $\mathrm{occur}(w_{13}, x) = 1$. From $w_{22} = w_{12}$ it follows that $\mathrm{occur}(w_1, x) > \mathrm{occur}(w_2, x)$. This is a contradiction to Property 8 of Definition 2.

– If $|r| > |u|$ and $rr \notin \mathrm{Suf}(w_2xu)$ and $r \notin \mathrm{Suf}(w_2xu)$ then: Let $w_{11}, w_{12}, w_{13} \in \Sigma_k^*$ be such that $w_1 = w_{11}w_{12}w_{13}$, $w_{12} = r$ and $w_{13}w_2xu = r$; see Figure below.

<table>
<tr><td>w_{11}</td><td>w_{12}</td><td>w_{13}</td><td>w_2</td><td>x</td><td>u</td></tr>
<tr><td></td><td>r</td><td colspan="4">r</td></tr>
</table>

It follows that

$$\mathrm{occur}(w_{12}, x) = \mathrm{occur}(w_{13}, x) + \mathrm{occur}(w_2, x) + \mathrm{occur}(xu, x).$$

This is a contradiction to Property 8 of Definition 2.

We proved that the assumption of existence of r, u leads to a contradiction. Thus we proved that for each prefix $u \in \mathrm{Prf}(t)$ we have that $w_1w_2xu \in \mathrm{L}_{k,\alpha}$. The proposition follows. □

We prove that if $(k, \alpha) \in \Upsilon$ then there is a right infinite α-power free word over Σ_{k-1}. In the introduction we showed that this observation could be deduced from Dejean's conjecture. Here additionally, to be able to address Problem 5 from the list of Restivo and Salemi, we present in the proof also examples of such words.

Lemma 2. *If $(k, \alpha) \in \Upsilon$ then the set $\mathrm{L}_{k-1,\alpha}^{\mathbb{N},R}$ is not empty.*

Proof. If $k = 3$ then $|\Sigma_{k-1}| = 2$. It is well known that the Thue Morse word is a right infinite 2^+-power free word over an alphabet with 2 letters [11]. It follows that the Thue Morse word is α-power free for each $\alpha > 2$.

If $k > 3$ then $|\Sigma_{k-1}| \geq 3$. It is well known that there are infinite 2-power free words over an alphabet with 3 letters [11]. Suppose $0, 1, 2 \in \Sigma_k$. An example is the fixed point of the morphism θ defined by $\theta(0) = 012$, $\theta(1) = 02$, and $\theta(2) = 1$ [11]. If an infinite word t is 2-power free then obviously t is α-power free and α^+-power free for each $\alpha \geq 2$.

This completes the proof. □

We define the sets of extendable words.

Definition 3. *Let $\mathrm{L} \subseteq \Sigma_k^*$. We define*

$$\mathrm{lext}(\mathrm{L}) = \{w \in \mathrm{L} \mid w \textit{ is left extendable in } \mathrm{L}\}$$

and

$$\mathrm{rext}(\mathrm{L}) = \{w \in \mathrm{L} \mid w \textit{ is right extendable in } \mathrm{L}\}.$$

If $u \in \mathrm{lext}(\mathrm{L})$ then let $\mathrm{lext}(u, \mathrm{L})$ be the set of all left infinite words $\bar{u}$ such that $\mathrm{Suf}(\bar{u}) \subseteq \mathrm{L}$ and $u \in \mathrm{Suf}(\bar{u})$. Analogously if $u \in \mathrm{rext}(\mathrm{L})$ then let $\mathrm{rext}(u, \mathrm{L})$ be the set of all right infinite words $\bar{u}$ such that $\mathrm{Prf}(\bar{u}) \subseteq \mathrm{L}$ and $u \in \mathrm{Prf}(\bar{u})$.

We show the sets $\text{lext}(u, \mathrm{L})$ and $\text{rext}(v, \mathrm{L})$ are nonempty for left extendable and right extendable words.

Lemma 3. *If* $\mathrm{L} \subseteq \Sigma_k^*$ *and* $u \in \text{lext}(\mathrm{L})$ *(resp.,* $v \in \text{rext}(\mathrm{L})$*) then* $\text{lext}(u, \mathrm{L}) \neq \emptyset$ *(resp.,* $\text{rext}(v, \mathrm{L}) \neq \emptyset$*).*

Proof. Realize that $u \in \text{lext}(\mathrm{L})$ (resp., $v \in \text{rext}(\mathrm{L})$) implies that there are infinitely many finite words in L having u as a suffix (resp., v as a prefix). Then the lemma follows from König's Infinity Lemma [4,8]. □

The next proposition proves that if $(k, \alpha) \in \Upsilon$, w is a right extendable α-power free word, $\bar{w}$ is a right infinite α-power free word having the letter x as a recurrent factor and having w as a prefix, and t is a right infinite α-power free word over $\Sigma_k \backslash \{x\}$, then there are finite words w_1, w_2, g such that the 5-tuple (w_1, w_2, x, g, t) is in the set $\Gamma(k, \alpha)$ and w is a prefix of $w_1 w_2 x g$.

Proposition 2. *If* $(k, \alpha) \in \Upsilon$*,* $w \in \text{rext}(\mathrm{L}_{k,\alpha})$*,* $\bar{w} \in \text{rext}(w, \mathrm{L}_{k,\alpha})$*,* $x \in \mathrm{F}_r(\bar{w}) \cap \Sigma_k$*,* $t \in \mathrm{L}_{k,\alpha}^{\mathbb{N},R}$*, and* $\text{occur}(t, x) = 0$ *then there are finite words* w_1, w_2, g *such that* $(w_1, w_2, x, g, t) \in \Gamma(k, \alpha)$ *and* $w \in \text{Prf}(w_1 w_2 x g)$*.*

Proof. Let $\omega = \mathrm{F}(\bar{w}) \cap \text{Prf}(xt)$ be the set of factors of $\bar{w}$ that are also prefixes of the word xt. Based on the size of the set ω we construct the words w_1, w_2, g and we show that $(w_1, w_2, x, g, t) \in \Gamma(k, \alpha)$ and $w_1 w_2 x g \in \text{Prf}(\bar{w}) \subseteq \mathrm{L}_{k,\alpha}$. The Properties 1, 2, 3, 4, 5, and 6 of Definition 2 are easy to verify. Hence we explicitly prove only properties 7 and 8 and that $w \in \text{Prf}(w_1 w_2 x g)$.

- If ω is an infinite set. It follows that $\text{Prf}(xt) = \omega$. Let $g \in \text{Prf}(t)$ be such that $|g| = |w|$; recall that t is infinite and hence such g exists. Let $w_2 \in \text{Prf}(\bar{w})$ be such that $w_2 x g \in \text{Prf}(\bar{w})$ and $\text{occur}(w_2 x g, x g) = 1$. Let $w_1 = \epsilon$.
  Property 7 of Definition 2 follows from $\text{occur}(w_2 x g, x g) = 1$. Property 8 of Definition 2 is obvious, because w_1 is the empty word. Since $|g| = |w|$ and $w \in \text{Prf}(\bar{w})$ we have that $w \in \text{Prf}(w_1 w_2 x g)$.
- If ω is a finite set. Let $\bar{\omega} = \omega \cap \mathrm{F}_r(\bar{w})$ be the set of prefixes of xt that are recurrent in $\bar{w}$. Since x is recurrent in $\bar{w}$ we have that $x \in \bar{\omega}$ and thus $\bar{\omega}$ is not empty. Let $g \in \text{Prf}(t)$ be such that xg is the longest element in $\bar{\omega}$. Let $w_1 \in \text{Prf}(w)$ be the shortest prefix of $\bar{w}$ such that if $u \in \omega \backslash \bar{\omega}$ is a non-recurrent prefix of xt in $\bar{w}$ then $\text{occur}(w_1, u) = \text{occur}(\bar{w}, u)$. Such w_1 obviously exists, because ω is a finite set and non-recurrent factors have only a finite number of occurrences. Let w_2 be the shortest factor of $\bar{w}$ such that $w_1 w_2 x g \in \text{Prf}(\bar{w})$, $\text{occur}(w_1, x) < \text{occur}(w_2, x)$, and $w \in \text{Prf}(w_1 w_2 x g)$. Since xg is recurrent in $\bar{w}$ and $w \in \text{Prf}(\bar{w})$ it is clear such w_2 exists.
  We show that Property 7 of Definition 2 holds. Let $y \in \Sigma_k$ be such that $gy \in \text{Prf}(t)$. Suppose that $\text{occur}(w_2 x g, x g y) > 0$. It would imply that xgy is recurrent in $\bar{w}$, since all occurrences of non-recurrent words from ω are in w_1. But we defined xg to be the longest recurrent word ω. Hence it is contradiction to our assumption that $\text{occur}(w_2 x g, x g y) > 0$.
  Property 8 of Definition 2 and $w \in \text{Prf}(w_1 w_2 x g)$ are obvious from the construction of w_2.

This completes the proof. $\square$

We define the *reversal* w^R of a finite or infinite word $w = \Sigma_k^* \cup \Sigma_k^{\mathbb{N}}$ as follows: If $w \in \Sigma_k^*$ and $w = w_1w_2\dots w_m$, where $w_i \in \Sigma_k$ and $1 \le i \le m$, then $w^R = w_mw_{m-1}\dots w_2w_1$. If $w \in \Sigma_k^{\mathbb{N},L}$ and $w = \dots w_2w_1$, where $w_i \in \Sigma_k$ and $i \in \mathbb{N}$, then $w^R = w_1w_2\dots \in \Sigma_k^{\mathbb{N},R}$. Analogously if $w \in \Sigma_k^{\mathbb{N},R}$ and $w = w_1w_2\dots$, where $w_i \in \Sigma_k$ and $i \in \mathbb{N}$, then $w^R = \dots w_2w_1 \in \Sigma_k^{\mathbb{N},L}$.

Proposition 1 allows one to construct a right infinite α-power free word with a given prefix. The next simple corollary shows that in the same way we can construct a left infinite α-power free word with a given suffix.

Corollary 1. *If $(k,\alpha) \in \Upsilon$, $w \in \text{lext}(\text{L}_{k,\alpha})$, $\bar{w} \in \text{lext}(w, \text{L}_{k,\alpha})$, $x \in \text{F}_r(\bar{w}) \cap \Sigma_k$, $t \in \text{L}_{k,\alpha}^{\mathbb{N},L}$, and $\text{occur}(t,x) = 0$ then there are finite words w_1, w_2, g such that $(w_1^R, w_2^R, x, g^R, t^R) \in \Gamma(k,\alpha)$, $w \in \text{Suf}(gxw_2w_1)$, and $txw_2w_1 \in \text{L}_{k,\alpha}^{\mathbb{N},L}$.*

Proof. Let $u \in \Sigma_k^* \cup \Sigma_k^{\mathbb{N}}$. Realize that $u \in \text{L}_{k,\alpha} \cup \text{L}_{k,\alpha}^{\mathbb{N}}$ if and only if $u^R \in \text{L}_{k,\alpha} \cup \text{L}_{k,\alpha}^{\mathbb{N}}$. Then the corollary follows from Proposition 1 and Proposition 2. $\square$

Given $k \in \mathbb{N}$ and a right infinite word $t \in \Sigma_k^{\mathbb{N},R}$, let $\Phi(t)$ be the set of all left infinite words $\bar{t} \in \Sigma_k^{\mathbb{N},L}$ such that $\text{F}(\bar{t}) \subseteq \text{F}_r(t)$. It means that all factors of $\bar{t} \in \Phi(t)$ are recurrent factors of t. We show that the set $\Phi(t)$ is not empty.

Lemma 4. *If $k \in \mathbb{N}$ and $t \in \Sigma_k^{\mathbb{N},R}$ then $\Phi(t) \neq \emptyset$.*

Proof. Since t is an infinite word, the set of recurrent factors of t is not empty. Let g be a recurrent nonempty factor of t; g may be a letter. Obviously there is $x \in \Sigma_k$ such that xg is also recurrent in t. This implies that the set $\{h \mid hg \in \text{F}_r(t)\}$ is infinite. The lemma follows from König's Infinity Lemma [4,8]. $\square$

The next lemma shows that if u is a right extendable α-power free word then for each letter x there is a right infinite α-power free word $\bar{u}$ such that x is recurrent in $\bar{u}$ and u is a prefix of $\bar{u}$.

Lemma 5. *If $(k,\alpha) \in \Upsilon$, $u \in \text{rext}(\text{L}_{k,\alpha})$, and $x \in \Sigma_k$ then there is $\bar{u} \in \text{rext}(u, \text{L}_{k,\alpha})$ such that $x \in \text{F}_r(\bar{u})$.*

Proof. Let $w \in \text{rext}(u, \text{L}_{k,\alpha})$; Lemma 3 implies that $\text{rext}(u, \text{L}_{k,\alpha})$ is not empty. If $x \in \text{F}_r(w)$ then we are done. Suppose that $x \notin \text{F}_r(w)$. Let $y \in \text{F}_r(w) \cap \Sigma_k$. Clearly $x \neq y$. Proposition 2 implies that there is $(w_1, w_2, y, g, t) \in \Gamma(k,\alpha)$ such that $u \in \text{Prf}(w_1w_2yg)$. The proof of Lemma 2 implies that we can choose t in such a way that x is recurrent in t. Then $w_1w_2yt \in \text{rext}(u, \text{L}_{k,\alpha})$ and $x \in \text{F}_r(w_1w_2yt)$. This completes the proof. $\square$

The next proposition shows that if u is left extendable and v is right extendable then there are finite words $\tilde{u}, \tilde{v}$, a letter x, a right infinite word t, and a left infinite word $\bar{t}$ such that $\tilde{u}xt, \bar{t}x\tilde{v}$ are infinite α-power free words, t has no occurrence of x, every factor of $\bar{t}$ is a recurrent factor in t, u is a prefix of $\tilde{u}xt$, and v is a suffix of $\bar{t}x\tilde{v}$.

Proposition 3. *If* $(k,\alpha)\in\Upsilon$, $u\in\mathrm{rext}(\mathrm{L}_{k,\alpha})$, *and* $v\in\mathrm{lext}(\mathrm{L}_{k,\alpha})$ *then there are* $\tilde{u},\tilde{v}\in\Sigma_k^*$, $x\in\Sigma_k$, $t\in\Sigma_k^{\mathbb{N},R}$, *and* $\bar{t}\in\Sigma_k^{\mathbb{N},L}$ *such that* $\tilde{u}xt\in\mathrm{L}_{k,\alpha}^{\mathbb{N},R}$, $\bar{t}x\tilde{v}\in\mathrm{L}_{k,\alpha}^{\mathbb{N},L}$, $\mathrm{occur}(t,x)=0$, $\mathrm{F}(\bar{t})\subseteq\mathrm{F}_r(t)$, $u\in\mathrm{Prf}(\tilde{u}xt)$, *and* $v\in\mathrm{Suf}(\bar{t}x\tilde{v})$.

Proof. Let $\bar{u}\in\mathrm{rext}(u,\mathrm{L}_{k,\alpha})$ and $\bar{v}\in\mathrm{lext}(v,\mathrm{L}_{k,\alpha})$ be such that $\mathrm{F}_r(\bar{u})\cap\mathrm{F}_r(\bar{v})\cap\Sigma_k\neq\emptyset$. Lemma 5 implies that such $\bar{u},\bar{v}$ exist. Let $x\in\mathrm{F}_r(\bar{u})\cap\mathrm{F}_r(\bar{v})\cap\Sigma_k$. It means that the letter x is recurrent in both $\bar{u}$ and $\bar{v}$.

Let t be a right infinite α-power free word over $\Sigma_k\setminus\{x\}$. Lemma 2 asserts that such t exists. Let $\bar{t}\in\Phi(t)$; Lemma 4 shows that $\Phi(t)\neq\emptyset$. It is easy to see that $\bar{t}\in\mathrm{L}_{k,\alpha}^{\mathbb{N},L}$, because $\mathrm{F}(\bar{t})\subseteq\mathrm{F}_r(t)$ and $t\in\mathrm{L}_{k,\alpha}^{\mathbb{N},R}$.

Proposition 2 and Corollary 1 imply that there are $u_1,u_2,g,v_1,v_2,\bar{g}\in\mathrm{L}_{k,\alpha}$ such that

- $(u_1,u_2,x,g,t)\in\Gamma(k,\alpha)$,
- $(v_1^R,v_2^R,x,\bar{g}^R,\bar{t}^R)\in\Gamma(k,\alpha)$,
- $u\in\mathrm{Prf}(u_1u_2xg)$, and
- $v^R\in\mathrm{Prf}(v_1^Rv_2^Rx\bar{g}^R)$; it follows that $v\in\mathrm{Suf}(\bar{g}xv_2v_1)$.

Proposition 1 implies that $u_1u_2xt, v_1^Rv_2^Rx\bar{t}^R\in\mathrm{L}_{k,\alpha}^{\mathbb{N},R}$. It follows that $\bar{t}xv_2v_1\in\mathrm{L}_{k,\alpha}^{\mathbb{N},L}$. Let $\tilde{u}=u_1u_2$ and $\tilde{v}=v_2v_1$. This completes the proof. □

The main theorem of the article shows that if u is a right extendable α-power free word and v is a left extendable α-power free word then there is a word w such that uwv is α-power free. The proof of the theorem shows also a construction of the word w.

Theorem 1. *If* $(k,\alpha)\in\Upsilon$, $u\in\mathrm{rext}(\mathrm{L}_{k,\alpha})$, *and* $v\in\mathrm{lext}(\mathrm{L}_{k,\alpha})$ *then there is* $w\in\mathrm{L}_{k,\alpha}$ *such that* $uwv\in\mathrm{L}_{k,\alpha}$.

Proof. Let $\tilde{u},\tilde{v},x,t,\bar{t}$ be as in Proposition 3. Let $p\in\mathrm{Suf}(\bar{t})$ be the shortest suffix such that $|p|>\max\{|\tilde{u}x|,|x\tilde{v}|,|u|,|v|\}$. Let $h\in\mathrm{Prf}(t)$ be the shortest prefix such that $hp\in\mathrm{Prf}(t)$ and $|h|>|p|$; such h exists, because p is a recurrent factor of t; see Proposition 3. We show that $\tilde{u}xhpx\tilde{v}\in\mathrm{L}_{k,\alpha}$.

We have that $\tilde{u}xhp\in\mathrm{L}_{k,\alpha}$, since $hp\in\mathrm{Prf}(t)$ and Proposition 3 states that $\tilde{u}xt\in\mathrm{L}_{k,\alpha}^{\mathbb{N},R}$. Lemma 1 implies that it suffices to show that there are no $g\in\mathrm{Prf}(\tilde{v})$ and no $r\in\Sigma_k^*\setminus\{\epsilon\}$ such that $rr\in\mathrm{Suf}(\tilde{u}xhpxg)$ and $|rr|>|xg|$. To get a contradiction, suppose there are such r,g. We distinguish the following cases.

- If $|r|\leq|xg|$ then $rr\in\mathrm{Suf}(pxg)$, because $|p|>|x\tilde{v}|$ and $xg\in\mathrm{Prf}(x\tilde{v})$. This is a contradiction, since $px\tilde{v}\in\mathrm{Suf}(\bar{t}x\tilde{v})$ and $\bar{t}x\tilde{v}\in\mathrm{L}_{k,\alpha}^{\mathbb{N},L}$; see Proposition 3.
- If $|r|>|xg|$ then $|r|\leq\frac{1}{2}|\tilde{u}xhpxg|$, otherwise rr cannot be a suffix of $\tilde{u}xhpxg$. Because $|h|>|p|>\max\{|\tilde{u}x|,|x\tilde{v}|\}$ we have that $r\in\mathrm{Suf}(hpxg)$. Since $\mathrm{occur}(hp,x)=0$, $|h|>|p|>|x\tilde{v}|$, and $xg\in\mathrm{Suf}(r)$ it follows that there are words h_1,h_2 such that $\tilde{u}xhpxg=\tilde{u}xh_1h_2pxg$, $r=h_2pxg$ and $r\in\mathrm{Suf}(\tilde{u}xh_1)$. It follows that $xg\in\mathrm{Suf}(\tilde{u}xh_1)$ and because $\mathrm{occur}(h_1,x)=0$ we have that $|h_1|\leq|g|$. Since $|p|>|\tilde{u}x|$ we get that $|h_2pxg|>|\tilde{u}xg|\geq|\tilde{u}xh_1|$; hence $|r|>|\tilde{u}xh_1|$. This is a contradiction.

We conclude that there is no word r and no prefix $g \in \mathrm{Prf}(\tilde{v})$ such that $rr \in \mathrm{Suf}(\tilde{u}xhpxg)$. Hence $\tilde{u}xhpx\tilde{v} \in \mathrm{L}_{k,\alpha}$. Due to the construction of p and h we have that $u \in \mathrm{Prf}(\tilde{u}xhpx\tilde{v})$ and $v \in \mathrm{Suf}(\tilde{u}xhpx\tilde{v})$. This completes the proof. $\square$

Acknowledgments. The author acknowledges support by the Czech Science Foundation grant GAČR 13-03538S and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/205/OHK4/3T/14.

References

1. Carpi, A.: On Dejean's conjecture over largealphabets. Theor. Comput. Sci. **385**, 137–151 (2007)
2. Currie, J., Rampersad, N.: A proof of Dejean's conjecture. Math. Comp. **80**, 1063–1070 (2011)
3. Dejean, F.: Sur un théorème de Thue. J. Comb. Theor. Series A **13**, 90–99 (1972)
4. König, D.: Sur les correspondances multivoques des ensembles. Fundamenta Math. **8**, 114–134 (1926)
5. Ollagnier, J.M.: Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters. Theor. Comput. Sci. **95**, 187–205 (1992)
6. Pansiot, J.-J.: A propos d'une conjecture de F. Dejean sur les répétitions dans les mots. Discrete Appl. Math. **7**, 297–311 (1984)
7. Petrova, E.A., Shur, A.M.: Transition property for cube-free words. In: van Bevern, R., Kucherov, G. (eds.) CSR 2019. LNCS, vol. 11532, pp. 311–324. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-19955-5_27
8. Rampersad, N.: Overlap-free words and generalizations. A thesis, University of Waterloo (2007)
9. Rao, M.: Last cases of Dejean's conjecture. Theor. Comput. Sci. **412**, 3010–3018 (2011). Combinatorics on Words (WORDS 2009)
10. Restivo, A., Salemi, S.: Some decision results on nonrepetitive words. In: Apostolico, A., Galil, Z. (eds.) Combinatorial Algorithms on Words, pp. 289–295. Springer, Heidelberg (1985). https://doi.org/10.1007/978-3-642-82456-2_20
11. Shallit, J., Shur, A.: Subword complexity and power avoidance. Theor. Comput. Sci. **792**, 96–116 (2019). Special issue in honor of the 70th birthday of Prof. Wojciech Rytter
12. Shur, A.M.: Two-sided bounds for the growth rates of power-free languages. In: Diekert, V., Nowotka, D. (eds.) DLT 2009. LNCS, vol. 5583, pp. 466–477. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02737-6_38
13. Shur, A.M.: Growth properties of power-free languages. In: Mauri, G., Leporati, A. (eds.) DLT 2011. LNCS, vol. 6795, pp. 28–43. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22321-1_3

Article [[Ru08]]: Palindromic Length of Words with Many Periodic Palindromes

Palindromic Length of Words with Many Periodic Palindromes

Josef Rukavicka(✉)

Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic
josef.rukavicka@seznam.cz

Abstract. The palindromic length $\mathrm{PL}(v)$ of a finite word v is the minimal number of palindromes whose concatenation is equal to v. In 2013, Frid, Puzynina, and Zamboni conjectured that: If w is an infinite word and k is an integer such that $\mathrm{PL}(u) \leq k$ for every factor u of w then w is ultimately periodic.

Suppose that w is an infinite word and k is an integer such $\mathrm{PL}(u) \leq k$ for every factor u of w. Let $\Omega(w,k)$ be the set of all factors u of w that have more than $\sqrt[k]{k^{-1}|u|}$ palindromic prefixes. We show that $\Omega(w,k)$ is an infinite set and we show that for each positive integer j there are palindromes a,b and a word $u \in \Omega(w,k)$ such that $(ab)^j$ is a factor of u and b is nonempty. Note that $(ab)^j$ is a periodic word and $(ab)^i a$ is a palindrome for each $i \leq j$. These results justify the following question: What is the palindromic length of a concatenation of a suffix of b and a periodic word $(ab)^j$ with "many" periodic palindromes?

It is known that if u,v are nonempty words then $|\mathrm{PL}(uv) - \mathrm{PL}(u)| \leq \mathrm{PL}(v)$. The main result of our article shows that if a,b are palindromes, b is nonempty, u is a nonempty suffix of b, $|ab|$ is the minimal period of aba, and j is a positive integer with $j \geq 3\mathrm{PL}(u)$ then $\mathrm{PL}(u(ab)^j) - \mathrm{PL}(u) \geq 0$.

1 Introduction

In 2013, Frid, Puzynina, and Zamboni introduced a *palindromic length* of a finite word [6]. Recall that the word $u = x_1x_2\dots x_n$ of length n is called a *palindrome* if $x_1x_2\dots x_n = x_n\dots x_2x_1$, where x_i are letters and $i \in \{1,2,\dots,n\}$. The palindromic length $\mathrm{PL}(u)$ of the word u is defined as the minimal number k such that $u = u_1u_2\dots u_k$ and u_j are palindromes, where $j \in \{1,2,\dots,k\}$; note that the palindromes u_j are not necessarily distinct. Let ϵ denote the empty word. We define that $\mathrm{PL}(\epsilon) = 0$.

In general, the factorization of a finite word into the minimal number of palindromes is not unique; for example $\mathrm{PL}(011001) = 3$ and the word 011001 can be factorized in two ways: $011001 = (0110)(0)(1) = (0)(1)(1001)$.

The authors of [6] conjectured that:

Conjecture 1. If w is an infinite word and P is an integer such that $\mathrm{PL}(u) \leq P$ for every factor u of w then w is ultimately periodic.

G. Jirásková and G. Pighizzini (Eds.): DCFS 2020, LNCS 12442, pp. 167–179, 2020.
https://doi.org/10.1007/978-3-030-62536-8_14

So far, Conjecture 1 remains open. We call an infinite word that satisfies the condition from Conjecture 1 a word with a *bounded palindromic length*. Note that there are infinite periodic words that do not have a bounded palindromic length; for example $(012)^\infty$. Hence the converse of Conjecture 1 does not hold.

In [6] the conjecture was proved for infinite words that are k-power free for some positive integer k. It follows that if w is an infinite word with a bounded palindromic length, then for each positive integer j there is a nonempty factor r such that r^j is a factor of w.

In [11], another variation of Conjecture 1 was considered:

Conjecture 2. Every aperiodic (not ultimately periodic) infinite word has prefixes of arbitrarily high palindromic length.

In [11], the author proved that Conjecture 1 and Conjecture 2 are equivalent. More precisely, it was proved that if every prefix of an infinite word w is a concatenation of at most n palindromes then every factor of w is a concatenation of at most $2n$ palindromes. It follows that Conjecture 2 remains also open.

In [7] Conjecture 1 and Conjecture 2 have been proved for all Sturmian words. The properties of the palindromic length of Sturmian words have been investigated also in [2]. In [1], the authors study the palindromic length of factors of fixed points of primitive morphisms. In [8], the lower bounds for the palindromic length of prefixes of infinite words can be found.

In [4], a left and right greedy palindromic length have been introduced as a variant to the palindromic length. It is shown that if the left (or right) greedy palindromic lengths of prefixes of an infinite word w is bounded, then w is ultimately periodic.

In addition, algorithms for computing the palindromic length were researched [3,5,10]. In [10], the authors present a linear time online algorithm for computing the palindromic length.

In the current paper we investigate infinite words with a bounded palindromic length. Let k be a positive integer, let w be an infinite word such that $k \geq \mathrm{PL}(t)$ for every factor t of w, and let $\Omega(w,k)$ be the set of all factors u of w that have more than $\sqrt[k]{k^{-1}|u|}$ palindromic prefixes. We show that $\Omega(w,k)$ is an infinite set and we show that for each positive integer j there are palindromes a,b and a word $u \in \Omega(w,k)$ such that $(ab)^j$ is a factor of u and b is nonempty. Note that $(ab)^j$ is a periodic word and $(ab)^i a$ is a palindrome for each $i \leq j$. In this sense we can consider that w has infinitely many periodic palindromes with an arbitrarily high exponent j.

The existence of infinitely many periodic palindromes in w is not surprising. It can be deduced also from the result in [6], which says, as mentioned above, that if w is an infinite word with a bounded palindromic length, then for each positive integer j there is a nonempty factor r such that r^j is a factor of w.

These results justify the following question: What is the palindromic length of a concatenation of a suffix of b and a periodic word $(ab)^j$ with "many" periodic palindromes?

It is known that if u,v are nonempty words then $|\mathrm{PL}(uv) - \mathrm{PL}(u)| \leq \mathrm{PL}(v)$ [11]. Less formally said, it means that by concatenating a word v to a word u the change of the palindromic length is at most equal to the palindromic length of v. The main result of our article shows that if a, b are palindromes, b is nonempty, u is a nonempty suffix of b, $|ab|$ is the minimal period of aba, and j is a positive integer with $j \geq 3\mathrm{PL}(u)$ then $\mathrm{PL}(u(ab)^j) - \mathrm{PL}(u) \geq 0$.

The results of our article should shed some light on infinite words for which Conjecture 1 and Conjecture 2 remain open. For the moment, for given factor u, we identified factors v such that $\mathrm{PL}(uv) - \mathrm{PL}(v) \geq 0$. The idea for the future development of this result is, for given $k \in \mathbb{N}$, to identify factors u, v such that $\mathrm{PL}(u) = k$ and $\mathrm{PL}(uv) - \mathrm{PL}(u) > 0$. The existence of such factors would, in consequence, allow us to prove the Conjecture 1 and Conjecture 2.

2 Preliminaries

Let $\mathbb{N}$ denote the set of all positive integers, let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ denote the set of all nonnegative integers, let $\mathbb{R}$ denote the set of all real numbers, and let $\mathbb{R}^+$ denote the set of all positive real numbers.

Let A denote a finite alphabet with $|\mathrm{A}| \geq 2$ letters. Let A^+ denote the set of all finite nonempty words over the alphabet A and let $\mathrm{A}^* = \mathrm{A}^+ \cup \{\epsilon\}$; recall that ϵ denotes the empty word. Let $\mathrm{A}^{\mathbb{N}}$ denote the set of all right infinite words.

Let $n \in \mathbb{N}$ and let $w = w_1 w_2 \dots w_n \in \mathrm{A}^*$ (or $w = w_1 w_2 \dots \in \mathrm{A}^{\mathbb{N}}$), where $w_i \in \mathrm{A}$ and $i \in \{1, 2, \dots, n\}$ (or $i \in \{1, 2, \dots\}$). We denote by $w[i, j] = w_i w_{i+1} \dots w_j$ the factor of w starting at position $i \in \mathbb{N}$ and ending at position $j \in \mathbb{N}$, where $i, j \in \mathbb{N}$ and $i \leq j \leq n$

We call the word $v \in \mathrm{A}^*$ a *factor* of the word $w \in \mathrm{A}^* \cup \mathrm{A}^{\mathbb{N}}$ if there are words $a \in \mathrm{A}^*$ and $b \in \mathrm{A}^* \cup \mathrm{A}^{\mathbb{N}}$ such that $w = avb$. Given a word $w \in \mathrm{A}^* \cup \mathrm{A}^{\mathbb{N}}$, we denote by $\mathrm{Fac}(w)$ the set of all factors of w. It follows that $\epsilon \in \mathrm{Fac}(w)$ and if $w \in \mathrm{A}^*$ then also $w \in \mathrm{Fac}(w)$.

We call the word $v \in \mathrm{A}^*$ a *prefix* of the word $w \in \mathrm{A}^* \cup \mathrm{A}^{\mathbb{N}}$ if there is $t \in \mathrm{A}^* \cup \mathrm{A}^{\mathbb{N}}$ such that $w = vt$. Given a word $w \in \mathrm{A}^* \cup \mathrm{A}^{\mathbb{N}}$, we denote by $\mathrm{Prf}(w)$ the set of all prefixes of w. It follows that $\epsilon \in \mathrm{Prf}(w)$ and if $w \in \mathrm{A}^*$ then also $w \in \mathrm{Prf}(w)$.

We call the word $v \in \mathrm{A}^*$ a *suffix* of the word $w \in \mathrm{A}^*$ if there is $t \in \mathrm{A}^*$ such that $w = tv$. Given a word $w \in \mathrm{A}^*$, we denote by $\mathrm{Suf}(w)$ the set of all suffixes of w. It follows that $\epsilon, w \in \mathrm{Suf}(w)$.

Let $w = w_1 w_2 \dots w_n \in \mathrm{A}^+$, where $w_i \in \mathrm{A}$ and $i \in \{1, 2, \dots, n\}$. Let w^R denote the *reversal* of the word $w \in \mathrm{A}^+$; it means $w^R = w_n w_{n-1} \dots w_2 w_1$. In addition we define that the reversal of the empty word is the empty word; formally $\epsilon^R = \epsilon$.

Realize that $w \in \mathrm{A}^*$ is a *palindrome* if and only if $w^R = w$. Let $\mathrm{Pal} \subset \mathrm{A}^*$ denote the set of all palindromes over the alphabet A. We define that $\epsilon \in \mathrm{Pal}$. Let $\mathrm{Pal}^+ = \mathrm{Pal} \setminus \{\epsilon\}$ be the set of all nonempty palindromes.

Given $w \in \mathrm{A}^* \cup \mathrm{A}^{\mathbb{N}}$, let $\mathrm{PalPrf}(w) = \mathrm{Pal} \cap \mathrm{Prf}(w)$ be the set of all palindromic prefixes of w.

Given $w \in \mathrm{A}^+$, let $\mathrm{MPF}(w)$ denote the set of all k-tuples of palindromes whose concatenation is equal to w and $k = \mathrm{PL}(w)$; formally

$$\mathrm{MPF}(w) = \{(t_1, t_2, \dots, t_k) \mid k = \mathrm{PL}(w) \text{ and } t_1 t_2 \dots t_k = w \text{ and } t_1, t_2, \dots, t_k \in \mathrm{Pal}^+\}.$$

We call a k-tuple $(t_1, t_2, \dots, t_k) \in \mathrm{MPF}(w)$ a *minimal palindromic factorization* of w.

Let $\mathbb{Q}$ denote the set of all rational numbers. We say that the word $w \in \mathrm{A}^+$ is a *periodic* word, if there are $\alpha \in \mathbb{Q}$, $r \in \mathrm{Prf}(w)\setminus\{\epsilon\}$, and $\bar{r} \in \mathrm{Prf}(r)\setminus\{r\}$ such that $\alpha > 1$, $w = rr\dots r\bar{r}$, and $\frac{|w|}{|r|} = \alpha$; note that $\bar{r}$ is uniquely determined by r. We write $w = r^\alpha$ and the period of w is equal to $|r|$. For example $12341 = (1234)^{\frac{5}{4}}$ and $12341234123 = (1234)^{\frac{11}{4}}$.

Given $w \in \mathrm{A}^+$, let

$$\mathrm{Period}(w) = \{(r, \alpha) \mid r^\alpha = w \text{ and } r \in \mathrm{Prf}(w) \setminus \{\epsilon\} \text{ and } \alpha \in \mathbb{Q} \text{ and } \alpha > 1\}.$$

The set $\mathrm{Period}(w)$ contains all couples (r, α) such that $r^\alpha = w$. Let

$$\mathrm{MinPer}(w) = \min\{|r| \mid (r, \alpha) \in \mathrm{Period}(w)\} \in \mathbb{N}.$$

The positive integer $\mathrm{MinPer}(w)$ is the *minimal period* of the word w. The word $w \in \mathrm{A}^+$ has a period $\delta \in \mathbb{Q}$ if there is a couple $(r, \alpha) \in \mathrm{Period}(w)$ such that $|r| = \delta$.

We will deal a lot with periodic palindromes. The two following known lemmas will be useful for us.

Lemma 1 *(see [9, Lemma 1]).* *Suppose p is a period of a nonempty palindrome w; then there are palindromes a and b such that $|ab| = p$, $b \neq \epsilon$, and $w = (ab)^* a$.*

Lemma 2 *(see [9, Lemma 2]).* *Suppose w is a palindrome and u is its proper suffix-palindrome or prefix-palindrome; then the number $|w| - |u|$ is a period of w.*

3 Periodic Palindromic Factors

We start the section with a definition of a set of real non-decreasing functions that diverge as n tends towards the infinity.

Let Λ denote the set of all functions ϕ such that

- $\phi(n) : \mathbb{N} \to \mathbb{R}$,
- $\phi(n) \leq \phi(n+1)$, and
- $\lim_{n\to\infty} \phi(n) = \infty$.

Let $k \in \mathbb{N}$, let $\tau(n,k) = \sqrt[k]{k^{-1}n} \in \Lambda$, let $w \in \mathrm{A}^{\mathbb{N}}$, and let

$$\Omega(w,k) = \{t \in \mathrm{Fac}(w) \mid |\mathrm{PalPrf}(t)| \geq \tau(|t|,k)\}.$$

The definition says that the set $\Omega(w,k)$ contains a factor t of w if the number of palindromic prefixes of t is larger than or equal to $\tau(|t|,k) = \sqrt[k]{k^{-1}|t|}$.

The next proposition asserts that if w is an infinite word with a bounded palindromic length, then the set of factors that have more than $\tau(n,k)$ palindromic prefixes is infinite, where n is the length of the factor in question and $k \geq \mathrm{PL}(t)$ for each factor t of w.

Proposition 1. *If $w \in \mathrm{A}^{\mathbb{N}}$, $k \in \mathbb{N}$ and $k \geq \max\{\mathrm{PL}(t) \mid t \in \mathrm{Fac}(w)\}$ then $|\Omega(w,k)| = \infty$.*

Proof. Suppose that $|\Omega(w,k)| < \infty$ and let

$$K = \max\{|\mathrm{PalPrf}(t)| \mid t \in \Omega(w,k)\}.$$

Less formally said, the value K is the maximal value from the set of numbers of palindromic prefixes of factors t of w that have more than $\tau(|t|,k)$ palindromic prefixes. Clearly $K < \infty$, because of the assumption $|\Omega(w,k)| < \infty$.

Let $p \in \mathrm{Prf}(w)$ be the shortest prefix of w such that $\tau(|p|,k) > K$. Since $\lim_{n\to\infty} \tau(n,k) = \infty$, it is clear that such prefix p exists.

To get a contradiction suppose that $|\mathrm{PalPrf}(t)| \geq \tau(|p|,k)$ for some $t \in \mathrm{Fac}(p)$. Since $\tau(|t|,k) \leq \tau(|p|,k)$ and thus $|\mathrm{PalPrf}(t)| \geq \tau(|t|,k)$, it follows that $t \in \Omega(w,k)$ and consequently $|\mathrm{PalPrf}(t)| \leq K$. It is a contradiction, because $K < \tau(|p|,k)$. Hence we have that

$$|\mathrm{PalPrf}(t)| < \tau(|p|,k) \text{ for each } t \in \mathrm{Fac}(p). \tag{1}$$

Let $n, j \in \mathbb{N}$ and let

$$\begin{aligned}\Theta(n,j) = \{(v_1, v_2, \ldots, v_j) \mid v_i \in \mathrm{Pal}^+ \text{ and } i \in \{1,2,\ldots,j\} \text{ and}\\ |v_1 v_2 \ldots v_j| \leq n \text{ and } v_1 v_2 \ldots v_j \in \mathrm{Prf}(w)\}.\end{aligned}$$

The set $\Theta(n,j)$ contains j-tuples of nonempty palindromes whose concatenation is of length smaller than or equal to n and also the concatenation is a prefix of w.

Thus from (1) we get that

$$|\Theta(|p|,j)| < (\tau(|p|,k))^j. \tag{2}$$

The Eq. (2) follows from the fact that each factor of p has at most $\tau(|p|,k)$ palindromic prefixes. In consequence there are at most $(\tau(|p|,k))^j$ of j-tuples of palindromes.

Let $\bar{\Theta}(|p|,j) = \bigcup_{j>0}^{k} \Theta(|p|,j)$. Since $\tau(n,k) \leq \tau(n+1,k)$ we have from (2) that

$$|\bar{\Theta}(|p|,k)| \leq k|\Theta(|p|,k)| < k(\tau(|p|,k))^k \leq k\left(\sqrt[k]{k^{-1}|p|}\right)^k = |p|. \tag{3}$$

The inequality (3) says that the number of prefixes of p having the form $v_1 v_2 \ldots v_j$, where $j \leq k$ and $v_i \in \mathrm{Pal}^+$ is smaller than the length of p. But p has $|p|$ nonempty prefixes. It is a contradiction. Since $\bigcup_{r\in\mathrm{Prf}(p)} \mathrm{MPF}(r) \subseteq \bar{\Theta}(|p|,k)$ we conclude that $\Omega(w,k)$ is an infinite set.

Remark 1. In the proof of Proposition 1, we used the idea that the number of prefixes of a word of length n that are a concatenation of at most k palindromes is smaller than n. This idea was used also in Theorem 1 in [6].

We show that if Σ is an infinite set of words r such that the number of nonempty palindromic prefixes of r grows more than $\ln |r|$ as $|r|$ tends towards infinity then for each positive integer j there are palindromes a, b and a word $t \in \Sigma$ such that $(ab)^j$ is a prefix of t and b is nonempty. Realize that $(ab)^j a$ is a palindrome for each $j \in \mathbb{N}_0$. This means that Σ contains infinitely many words that have a periodic palindromic prefix of arbitrarily high exponent j.

Proposition 2. *If $\Sigma \subseteq \mathrm{A}^*$, $|\Sigma| = \infty$, $\phi(n) \in \Lambda$, $\lim_{n\to\infty} (\phi(n) - \ln n) = \infty$, and $|\mathrm{PalPrf}(t) \setminus \{\epsilon\}| \geq \phi(|t|)$ for each $t \in \Sigma$ then for each $j \in \mathbb{N}$ there are palindromes $a \in \mathrm{Pal}$, $b \in \mathrm{Pal}^+$ and a word $t \in \Sigma$ such that $(ab)^j \in \mathrm{Prf}(t)$.*

Proof. Given $t \in \Sigma$, let $\mu(t, i)$ be the lengths of all palindromic prefixes of t such that $\mu(t, 1) = 1$ (a letter is a palindrome) and $\mu(t, i) < \mu(t, i + 1)$, where $i \in \{1, 2, \dots, h_t\}$. For example if $t = 0100010111$, then $\mu(t, 1) = |0| = 1$, $\mu(t, 2) = |010| = 3$, $\mu(t, 3) = |0100010| = 7$. Let $h_t = |\mathrm{PalPrf}(t) \setminus \{\epsilon\}|$; the integer h_t is the number of nonempty palindromic prefixes of t. Let $i \in \{1, 2, \dots, h_t - 1\}$. It is clear that

$$\mu(t, i + 1) = \mu(t, i)\frac{\mu(t, i + 1)}{\mu(t, i)}. \tag{4}$$

From (4) we have that

$$\frac{\mu(t, h_t)}{\mu(t, h_t - 1)}\frac{\mu(t, h_t - 1)}{\mu(t, h_t - 2)}\frac{\mu(t, h_t - 2)}{\mu(t, h_t - 3)}\cdots\frac{\mu(t, 2)}{\mu(t, 1)} = \mu(t, h_t) \leq |t|. \tag{5}$$

Suppose that there is $\alpha \in \mathbb{R}$ such that $\alpha > 1$ and for each $t \in \Sigma$ and for each $i \in \{1, 2, \dots, h_t - 1\}$ we have that $\frac{\mu(t,i+1)}{\mu(t,i)} \geq \alpha$. It follows from (5) that

$$\alpha^{h_t - 1} \leq h_t \leq |t|. \tag{6}$$

Let $c = \frac{1}{\ln \alpha} \in \mathbb{R}^+$. Then $|t| = \alpha^{c \ln |t|}$. Since $h_t \geq \phi(|t|)$ we get that

$$\frac{\alpha^{h_t - 1}}{|t|} \geq \frac{\alpha^{\phi(|t|)-1}}{|t|} = \frac{\alpha^{\phi(|t|)-1}}{\alpha^{c \ln |t|}} = \alpha^{\phi(|t|)-1-c \ln |t|}. \tag{7}$$

Because $\lim_{n\to\infty} (\phi(n) - \ln n) = \infty$ the Eq. (7) implies that there is n_0 such that for each $t \in \Sigma$ with $|t| > n_0$ we have that

$$\frac{\alpha^{h_t - 1}}{|t|} \geq \alpha^{\phi(|t|)-1-c \ln |t|} > 1. \tag{8}$$

From (6) and (8) we have that $\alpha^{h_t - 1} \leq |t|$ and $\frac{\alpha^{h_t-1}}{|t|} > 1$, which is a contradiction. We conclude there is no such α. In consequence, for each $\beta \in \mathbb{R}^+$ with $\beta > 1$ there is $t \in \Sigma$ and $i \in \{1, 2, \dots, h_t - 1\}$ such that $\frac{\mu(t,i+1)}{\mu(t,i)} \leq \beta$.

Let $j \in \mathbb{N}$, let

$$\gamma \leq \frac{1}{j} + 1 \in \mathbb{R}^+, \tag{9}$$

let $t \in \Sigma$, and $i \in \{1, 2, \dots, h_t\}$ be such that $\frac{\mu(t,i+1)}{\mu(t,i)} \leq \gamma$. Let $\delta = \frac{\mu(t,i+1)}{\mu(t,i)} \leq \gamma$. Let $u, v \in \mathrm{Prf}(t)$ be such that $|u| = \mu(t,i)$ and $|v| = \mu(t,i+1)$. Then v is a periodic palindrome with a period $|v| - |u| = \mu(t,i+1) - \mu(t,i) = \mu(t,i)\delta - \mu(t,i) = \mu(t,i)(\delta - 1)$; see Lemma 2. Lemma 1 implies that there are $a \in \mathrm{Pal}$ and $b \in \mathrm{Pal}^+$ such that $(ab)^k a = v$ for some $k \in \mathbb{N}$. From Lemma 1 we have also that $|ab|$ is the period of v. Thus

$$|ab| = \mu(t,i)(\delta - 1) \leq \mu(t,i)(\gamma - 1). \tag{10}$$

From (9) and (10) it follows that

$$|ab| \leq \mu(t,i)(\gamma - 1) \leq \mu(t,i)\frac{1}{j}. \tag{11}$$

Note that $v = (ab)^k a$ and $u \in \mathrm{Prf}((ab)^k)$. Since $\mu(t,i) = |u|$ we get that $\frac{\mu(t,i)}{|ab|} \leq k$. From (11) we have that

$$j \leq \frac{\mu(t,i)}{|ab|} \leq k.$$

Thus for arbitrary $j \in \mathbb{N}$ we found t, a, b, k such that $(ab)^k \in \mathrm{Prf}(t)$ and $j \leq k$. The proposition follows.

A corollary of Proposition 1 and Proposition 2 says that if w is an infinite word with a bounded palindromic length then for each positive integer j there are palindromes a, b such that $(ab)^j$ is a factor of w and ab is a nonempty word.

Corollary 1. *If* $w \in \mathrm{A}^{\mathbb{N}}$, $k \in \mathbb{N}$, *and* $k \geq \max\{\mathrm{PL}(t) \mid t \in \mathrm{Fac}(w)\}$ *then for each* $j \in \mathbb{N}$ *there are* $a \in \mathrm{Pal}$ *and* $b \in \mathrm{Pal}^+$ *such that* $(ab)^j \in \mathrm{Fac}(w)$.

Proof. Just take $\Sigma = \Omega(w,k)$. Obviously $\lim_{n\to\infty}(\tau(n,k) - \ln n) = \infty$. Then Proposition 2 implies the corollary.

4 Palindromic Length of Concatenation

In this section we present some known results about the palindromic length of concatenation of two words.

The first lemma shows the very basic property of the palindromic length that the palindromic length of concatenation of two words x and y is smaller than or equal to the sum of palindromic length of x and y. We omit the proof.

Lemma 3. *If* $x, y \in \mathrm{A}^*$ *then* $\mathrm{PL}(xy) \leq \mathrm{PL}(x) + \mathrm{PL}(y)$.

An another basic property says that if $(t_1, t_2, \ldots, t_k) \in \mathrm{MPF}(w)$ is a minimal palindromic factorization of the word w then the palindromic length of the factor $t_i t_{i+1} \ldots t_j$ is equal to $j - i + 1$ for each $i, j \in \{1, 2, \ldots, k\}$ and $i \leq j$. We omit the proof.

Lemma 4. *If* $w \in \mathrm{A}^+$, $k = \mathrm{PL}(w)$, *and* $(t_1, t_2, \ldots, t_k) \in \mathrm{MPF}(w)$ *then for each* $i, j \in \{1, 2, \ldots, k\}$ *with* $i \leq j$ *we have that* $\mathrm{PL}(t_i t_{i+1} \ldots t_j) = j - i + 1$.

The following result has been proved in [11]. It says that if x, y are words then the palindromic length of y is the maximal absolute difference of palindromic lengths of x and xy; i.e. $|\mathrm{PL}(x) - \mathrm{PL}(xy)| \leq \mathrm{PL}(y)$.

Lemma 5 *(see [11, Lemma 6])***.** *If* $x, y \in \mathrm{A}^*$ *then*

- $\mathrm{PL}(y) \leq \mathrm{PL}(x) + \mathrm{PL}(xy)$ *and*
- $\mathrm{PL}(x) \leq \mathrm{PL}(y) + \mathrm{PL}(xy)$.

We have two following immediate corollaries of Lemma 5.

Corollary 2. *If* $x, y \in \mathrm{A}^*$ *and* $y \in \mathrm{Pal}$ *then* $|\mathrm{PL}(xy) - \mathrm{PL}(x)| \leq 1$.

Proof. It is enough to consider y in Lemma 5 to be a palindrome. Thus we have $\mathrm{PL}(y) = 1$ if $y \neq \epsilon$ or $\mathrm{PL}(y) = 0$ if $y = \epsilon$. The corollary follows.

Corollary 3. *If* $x, y \in \mathrm{A}^*$ *and* $xy \in \mathrm{Pal}$ *then* $|\mathrm{PL}(x) - \mathrm{PL}(y)| \leq 1$.

Proof. If $x = y^R$ then $\mathrm{PL}(x) - \mathrm{PL}(y) = 0$, because clearly $\mathrm{PL}(y) = \mathrm{PL}(y^R)$. Suppose that $x \neq y^R$. It follows that $|x| \neq |y|$, since $xy \in \mathrm{Pal}$. Without loss of generality suppose that $|x| > |y|$. Let $\bar{x}$ be such that $x = y^R \bar{x}$. Then $xy = y^R \bar{x} y$. Thus $\bar{x} \in \mathrm{Pal}^+$. Corollary 2 implies that $|\mathrm{PL}(y^R \bar{x}) - \mathrm{PL}(y)| \leq 1$. The corollary follows.

5 Concatenation of Periodic Palindromes

To simplify the notation of the next two lemmas and the theorem we define an auxiliary set Δ. Let Δ be the set of all 4-tuples (u, d, v, n) such that

- $d \in \mathrm{Pal}^+$,
- $v \in \mathrm{Pal}$,
- $u \in \mathrm{Suf}(d) \setminus \{\epsilon\}$,
- $n \in \mathbb{N}$,
- $|dv| = \mathrm{MinPer}(dvd)$, and
- $n \geq 3\mathrm{PL}(u)$.

Remark 2. The set Δ contains all 4-tuples (u, v, d, n) such that d is a nonempty palindrome, v is a palindrome (possibly empty), u is a nonempty suffix of d, $|dv|$ is the minimal period of the word dvd, and n is a positive integer such that $n \geq 3\mathrm{PL}(u)$. It follows that $n \geq 3$, since u is nonempty and thus $\mathrm{PL}(u) \geq 1$.

Lemma 6. *If* $(u, v, d, n) \in \Delta$*,* $r \in \mathrm{Fac}(u(vd)^n)$*, and* $|r| \geq 3|vd|$ *then* $dvd \in \mathrm{Fac}(r)$*.*

Proof. Let $\bar{w} = u(vd)^n$, let $p \in \mathrm{Prf}(r)$ be such that $|p| = 3|vd|$, and let $\bar{i}, \bar{j} \in \{1, 2, \ldots, |\bar{w}|\}$ be such that $p = \bar{w}[\bar{i}, \bar{j}]$. Let $\bar{u} \in \mathrm{Prf}(d)$ be such that $d = \bar{u}u$. Note that $|uv\bar{u}| = |vd|$ and thus $(uv\bar{u}, \beta) \in \mathrm{Period}(\bar{w})$, where $\beta = \frac{|\bar{w}|}{|uv\bar{u}|} > 1$.

Let $k \in \mathbb{N}_0$ and $w \in \mathrm{Suf}(\bar{w})$ be such that $\bar{w} = (uv\bar{u})^k w$, $\bar{i} > |(uv\bar{u})^k|$, and $\bar{i} \leq |(uv\bar{u})^{k+1}|$. Obviously such k and w exist. Let $i = \bar{i} - k|uv\bar{u}|$ and $j = \bar{j} - k|uv\bar{u}|$. It is easy to see that $p = w[i, j]$.

We distinguish:

- If $i \in \{1, 2, \ldots, |u|\}$ then $p = tvdvdv\bar{t}$ for some $t \in \mathrm{Suf}(u)$ and for $\bar{t}$ such that $d = \bar{t}t$.
- If $i \in \{|u|+1, |u|+2, \ldots, |uv|\}$ then $p = tdvdvd\bar{t}$ for some $t \in \mathrm{Suf}(v)$ and for $\bar{t}$ such that $v = \bar{t}t$.
- If $i \in \{|uv|+1, |uv|+2, \ldots, |uv|+|\bar{u}|\}$ then $p = tvdvdv\bar{t}$ for some $t \in \mathrm{Suf}(d)$ and for $\bar{t}$ such that $d = \bar{t}t$.

In all three cases one can see that $dvd \in \mathrm{Fac}(p)$. It is easy to see that if $dvd \in \mathrm{Fac}(p)$ then $dvd \in \mathrm{Fac}(r)$ for each $r \in \mathrm{Fac}(w)$ with $p \in \mathrm{Prf}(r)$. The lemma follows.

Remark 3. Note in the previous proof that with the condition $|r| \geq |(vd)^2|$ it would be possible that $dvd \notin \mathrm{Fac}(p)$. In the cases 1 and 3 we would have $p = tvdv\bar{t}$. That is why the condition $|r| \geq |(vd)^3|$ is necessary. For this reason in the definition of Δ we state that $n \geq 3\mathrm{PL}(u)$.

The next lemma shows that if $(u, v, d, n) \in \Delta$, k is the palindromic length of u, and $(t_1, t_2, \ldots, t_k) \in \mathrm{MPF}(u(vd)^n)$ is a minimal palindromic factorization of $u(vd)^n$ then there is $j \in \{1, 2, \ldots, k\}$ such that t_j is a palindrome having the factor dvd in the "center" of t_j; formally $t_j = pd(vd)^\gamma p^R$ for some positive integer γ and for some proper suffix p of dv.

Lemma 7. *If* $(u, v, d, n) \in \Delta$*,* $w = u(vd)^n$*,* $k = \mathrm{PL}(w)$*, and* $(t_1, t_2, \ldots, t_k) \in \mathrm{MPF}(w)$ *then there are* $j \in \{1, 2 \ldots, k\}$*,* $p \in \mathrm{Suf}(dv) \setminus \{dv\}$*, and* $\gamma \in \mathbb{N}$ *such that* $t_j = pd(vd)^\gamma p^R$*.*

Proof. Suppose that $|t_i| < 3|vd|$ for each $i \in \{1, 2, \ldots, k\}$. It follows that

$$|t_1 t_2 \ldots t_k| < 3k|vd|.$$

Since $u(vd)^n = t_1 t_2 \ldots t_k$ and $n \geq 3k \geq 3$ it is a contradiction. It follows that there is j such that $|t_j| \geq |(vd)^3|$. Lemma 6 asserts that $dvd \in \mathrm{Fac}(t_j)$. Then clearly there are $\gamma \in \mathbb{N}$ and $p_1, p_2 \in \mathrm{A}^*$ such that $p_1 \in \mathrm{Suf}(dv) \setminus \{dv\}$, $p_2 \in \mathrm{Prf}(vd) \setminus \{vd\}$, and $t_j = p_1 d(vd)^\gamma p_2$.

To get a contradiction suppose that $p_1 \neq p_2^R$. Without loss of generality suppose that $|p_1| > |p_2|$. It follows that $p_2 \in \mathrm{Prf}(p_1^R)$. Obviously $p_1 d(vd)^\gamma p_1^R \in$

Pal. Thus we have two palindromes $p_1d(vd)^\gamma p_1^R$ and $p_1d(vd)^\gamma p_2$. Lemma 2 implies that $p_1d(vd)^\gamma p_1^R$ is periodic with a period

$$\delta = |p_1d(vd)^\gamma p_1^R| - |p_1d(vd)^\gamma p_2| = |p_1| - |p_2|.$$

Clearly $\delta < |dv|$. This is a contradiction to the condition $|dv| = \mathrm{MinPer}(dvd)$, see Definition of Δ. We conclude that $p_1 = p_2^R$. The lemma follows.

We step to the main theorem of the article.

Theorem 1. *If* $(u, v, d, n) \in \Delta$, $m = \mathrm{PL}(u)$, *and* $w = u(vd)^n$ *then* $\mathrm{PL}(w) \geq m$.

Proof. Let $(t_1, t_2, \ldots, t_k) \in \mathrm{MPF}(w)$. Lemma 7 asserts that there are $\gamma \in \mathbb{N}$, $j \in \{1, 2, \ldots, k\}$, and $p \in \mathrm{Suf}(dv) \setminus \{dv\}$ such that $t_j = pd(vd)^\gamma p^R$.

Let $a \in \mathrm{Prf}(w)$ and $b \in \mathrm{Suf}(w)$ be such that $w = at_jb$. Realize that $a = t_1t_2 \ldots t_{j-1}$ and $b = t_{j+1}t_{j+2} \ldots t_k$. Note that a or b can be the empty word; then $j = 1$ or $j = k$ respectively. Lemma 4 implies that

$$\begin{aligned}\mathrm{PL}(w) = \mathrm{PL}(t_1t_2 \ldots t_{j-1}) + \mathrm{PL}(t_j) + \mathrm{PL}(t_{j+1}t_{j+2} \ldots t_k) = \\ \mathrm{PL}(a) + \mathrm{PL}(t_j) + \mathrm{PL}(b).\end{aligned} \tag{12}$$

We distinguish three distinct cases.

1. $u \notin \mathrm{Prf}(a)$: This case is depicted in Table 1. Let $u_2 \in \mathrm{Suf}(u)$ be such that $u = au_2$. Let $\bar{p} \in \mathrm{Suf}(d)$ be such that $\bar{p}u_2 = d$. It follows that $u_2^R\bar{p}^R = d$ and $p^R\bar{p}^R = vd$.
   Then we have that $u_2^Rb = u_2^R\bar{p}^R(vd)^\beta = d(vd)^\beta \in \mathrm{Pal}^+$ for some $\beta \in \mathbb{N}_0$. Hence $\mathrm{PL}(u_2^R\bar{p}^R(vd)^\beta) = 1$. In consequence $\mathrm{PL}(u_2) \geq \mathrm{PL}(b) - 1$ and

   $$\mathrm{PL}(b) \geq \mathrm{PL}(u_2) - 1, \tag{13}$$

   since $\mathrm{PL}(u_2^R) = \mathrm{PL}(u_2)$ and $u_2^Rb \in \mathrm{Pal}^+$; see Corollary 3.
   Lemma 3 implies that

   $$\mathrm{PL}(a) + \mathrm{PL}(u_2) \geq \mathrm{PL}(u). \tag{14}$$

   From (12), (13), and (14) we have that

   $$\mathrm{PL}(w) = \mathrm{PL}(a) + \mathrm{PL}(t_j) + \mathrm{PL}(b) \geq \mathrm{PL}(a) + 1 + \mathrm{PL}(u_2) - 1 \geq \mathrm{PL}(u).$$

Table 1. Case 1: The structure of the word w with $u \notin \mathrm{Prf}(a)$.

a	t_j					b	
a	p		$d(vd)^\gamma$	p^R		$\bar{p}^R$	$(vd)^\beta$
a	u_2	v	$d(vd)^\gamma$	v	u_2^R	$\bar{p}^R$	$(vd)^\beta$
u		$(vd)^{\gamma+1}$		v	d		$(vd)^\beta$

2. $u \in \text{Prf}(a)$ and $p \in \text{Suf}(v)$: This case is depicted in Table 2. Let $\bar{p} \in \text{Prf}(v)$ be such that $\bar{p}p = v$. Note that if $p = v$ then $\bar{p} = \epsilon$, and if $p = \epsilon$ then $\bar{p} = v$. It is easy to verify that $b = \bar{p}^R d(vd)^\beta$ for some $\beta \in \mathbb{N}_0$ and $a = u(vd)^\alpha \bar{p}$ for some $\alpha \in \mathbb{N}_0$.
Let $\bar{a}$ be such that $a = u\bar{a}$. We have that $\bar{a} = (vd)^\alpha \bar{p}$ and $b = \bar{p}^R d(vd)^\beta$. It follows that either $\bar{a} = b^R d(vd)^\delta$ or $b = \bar{a}^R d(vd)^\delta$ for some $\delta \in \mathbb{N}_0$.
Since $d(vd)^\delta \in \text{Pal}$, Corollary 2 implies that
$$|\text{PL}(\bar{a}) - \text{PL}(b)| \leq 1. \tag{15}$$
It follows from Lemma 5 that
$$\text{PL}(a) + \text{PL}(\bar{a}) \geq \text{PL}(u). \tag{16}$$
From (12), (15), and (16) we have that
$$\text{PL}(w) = \text{PL}(a) + \text{PL}(t_j) + \text{PL}(b) \geq \text{PL}(a) + 1 + \text{PL}(\bar{a}) - 1 \geq \text{PL}(u).$$

Table 2. Case 2: The structure of the word w with $u \in \text{Prf}(a)$ and $p \in \text{Suf}(v)$.

a			t_j			b	
u	$(vd)^\alpha$	$\bar{p}$	p	$d(vd)^\gamma$	p^R	$\bar{p}^R$	$d(vd)^\beta$
	$\bar{a}$				v		

3. $u \in \text{Prf}(a)$ and $p \notin \text{Suf}(v)$: This case is depicted in Table 3. Since $p \in \text{Suf}(vd) \setminus \{vd\}$ and $p \notin \text{Suf}(v)$ it follows that $p \in \text{Suf}(dv) \setminus (\text{Suf}(v) \cup \{dv\})$.

Table 3. Case 3: The structure of the word w with $u \in \text{Prf}(a)$ and $p \notin \text{Suf}(v)$.

a			t_j			b	
u	$v(dv)^\alpha$	$\bar{p}$	p	$d(vd)^\gamma$	p^R	$\bar{p}^R$	$(vd)^\beta$
	$\bar{a}$				vd		

Let $\bar{p} \in \text{Prf}(d)$ be such that $\bar{p}p = dv$ and consequently $p^R \bar{p}^R = vd$. Then $a = u(vd)^\alpha \bar{p}$ for some $\alpha \in \mathbb{N}_0$ and $b = \bar{p}(vd)^\beta$ for some $\beta \in \mathbb{N}_0$.
Let $\bar{a}$ be such that $a = u\bar{a}$. We have that $\bar{a} = v(dv)^\alpha \bar{p}$. It follows that either $\bar{a} = b^R (vd)^\delta v$ or $b = \bar{a}^R (vd)^\delta v$ for some $\delta \in \mathbb{N}_0$.
The rest of the proof of Case 3 is analogue to Case 2: Since $v(dv)^\delta \in \text{Pal}$, Corollary 2 implies that
$$|\text{PL}(\bar{a}) - \text{PL}(b)| \leq 1. \tag{17}$$

It follows from Lemma 5 that

$$\mathrm{PL}(a) + \mathrm{PL}(\bar{a}) \geq \mathrm{PL}(u). \tag{18}$$

From (12), (17), and (18) we have that

$$\mathrm{PL}(w) = \mathrm{PL}(a) + \mathrm{PL}(t_j) + \mathrm{PL}(b) \geq \mathrm{PL}(a) + 1 + \mathrm{PL}(\bar{a}) - 1 \geq \mathrm{PL}(u).$$

We proved for each case that $\mathrm{PL}(w) \geq \mathrm{PL}(u)$. Since obviously for each u and each p one of the three cases applies, this completes the proof.

Acknowledgments. This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS20/183/OHK4/3T/14.

References

1. Ambrož, P., Kadlec, O., Masáková, Z., Pelantová, E.: Palindromic length of words and morphisms in class P. Theor. Comput. Sci. **780**, 74–83 (2019). https://doi.org/10.1016/j.tcs.2019.02.024
2. Ambrož, P., Pelantová, E.: On palindromic length of Sturmian sequences. In: Hofman, P., Skrzypczak, M. (eds.) Developments in Language Theory, pp. 244–250. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-24886-4_18
3. Borozdin, K., Kosolobov, D., Rubinchik, M., Shur, A.M.: Palindromic length in linear time. In: Kärkkäinen, J., Radoszewski, J., Rytter, W. (eds.) 28th Annual Symposium on Combinatorial Pattern Matching (CPM 2017). Leibniz International Proceedings in Informatics (LIPIcs), vol. 78, pp. 23:1–23:12. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2017). https://doi.org/10.4230/LIPIcs.CPM.2017.23
4. Bucci, M., Richomme, G.: Greedy palindromic lengths. Int. J. Found. Comput. Sci. **29**(03), 331–356 (2018). https://doi.org/10.1142/S0129054118500077
5. Fici, G., Gagie, T., Kräkkäinen, J., Kempa, D.: A subquadratic algorithm for minimum palindromic factorization. J. Discrete Algorithms **28**, 41–48 (2014). https://doi.org/10.1016/j.jda.2014.08.001. stringMasters 2012 & 2013 Special Issue (Volume 1)
6. Frid, A., Puzynina, S., Zamboni, L.: On palindromic factorization of words. Adv. Appl. Math. **50**(5), 737–748 (2013). https://doi.org/10.1016/j.aam.2013.01.002
7. Frid, A.E.: Sturmian numeration systems and decompositions to palindromes. Eur. J. Comb. **71**, 202–212 (2018). https://doi.org/10.1016/j.ejc.2018.04.003
8. Frid, A.E.: First lower bounds for palindromic length. In: Hofman, P., Skrzypczak, M. (eds.) Developments in Language Theory, pp. 234–243. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-24886-4_17
9. Kosolobov, D., Rubinchik, M., Shur, A.M.: Pal^k is linear recognizable online. In: Italiano, G.F., Margaria-Steffen, T., Pokorný, J., Quisquater, J.J., Wattenhofer, R. (eds.) SOFSEM 2015: Theory and Practice of Computer Science, pp. 289–301. Springer, Berlin Heidelberg, Berlin, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46078-8_24

10. Rubinchik, M., Shur, A.M.: EERTREE: an efficient data structure for processing palindromes in strings. In: Lipták, Z., Smyth, W.F. (eds.) IWOCA 2015. LNCS, vol. 9538, pp. 321–333. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29516-9_27
11. Saarela, A.: Palindromic length in free monoids and free groups. In: Brlek, S., Dolce, F., Reutenauer, C., Vandomme, É. (eds.) WORDS 2017. LNCS, vol. 10432, pp. 203–213. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66396-8_19

Article [[Ru09]]: Dissecting Power of a Finite Intersection of Context Free Languages

Dissecting Power of a Finite Intersection of Context Free Languages

Josef Rukavicka*

August 31, 2020
Mathematics Subject Classification: 68Q45

Abstract

Let $\exp^{k,\alpha}$ denote a tetration function defined as follows: $\exp^{1,\alpha} = 2^{\alpha}$ and $\exp^{k+1,\alpha} = 2^{\exp^{k,\alpha}}$, where k, α are positive integers. Let Δ_n denote an alphabet with n letters. If $L \subseteq \Delta_n^*$ is an infinite language such that for each $u \in L$ there is $v \in L$ with $|u| < |v| \leq \exp^{k,\alpha} |u|$ then we call L a language with the *growth bounded by* (k, α)-tetration.

Given two infinite languages $L_1, L_2 \in \Delta_n^*$, we say that L_1 *dissects* L_2 if $|L_1 \cap L_2| = \infty$ and $|(\Delta_n^* \setminus L_1) \cap L_2| = \infty$.

Given a context free language L, let $\kappa(L)$ denote the size of the smallest context free grammar G that generates L. We define the size of a grammar to be the total number of symbols on the right sides of all production rules.

Given positive integers n, k with $k \geq 2$, we show that there are context free languages $L_1, L_2, \dots, L_{3k-3} \subseteq \Delta_n^*$ with $\kappa(L_i) \leq 40k$ such that if α is a positive integer and $L \subseteq \Delta_n^*$ is an infinite language with the growth bounded by (k, α)-tetration then there is a regular language M such that $M \cap \left(\bigcap_{i=1}^{3k-3} L_i\right)$ dissects L and the minimal deterministic finite automaton accepting M has at most $k + \alpha + 3$ states.

*Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague (josef.rukavicka@seznam.cz).

1 Introduction

In the theory of formal languages, the regular and the context free languages constitute a fundamental concept that attracted a lot of attention in the past several decades. Recall that every regular language is accepted by some deterministic finite automaton and every context free language is accepted by some pushdown automaton.

In contrast to regular languages, the context free languages are closed neither under intersection nor under complement. The intersection of context free languages have been systematically studied in [4, 6, 9, 10, 11]. Let CFL_k denote the family of all languages such that for each $L \in \mathrm{CFL}_k$ there are k context free languages $L_1, L_2, \dots, L_k$ with $L = \bigcap_{i=1}^{k} L_i$. For each k, it has been shown that there is a language $L \in \mathrm{CFL}_{k+1}$ such that $L \notin \mathrm{CFL}_k$. Thus the k-intersections of context free languages form an infinite hierarchy in the family of all formal languages lying between context free and context sensitive languages [6].

One of the topics in the theory of formal languages that has been studied is the dissection of infinite languages. Let Δ_n be an alphabet with n letters, and let $L_1, L_2 \subseteq \Delta_n^*$ be infinite languages. We say that L_1 *dissects* L_2 if $|L_1 \cap L_2| = \infty$ and $|(\Delta_n^* \setminus L_1) \cap L_2| = \infty$. Let $\mathbb{C}$ be a family of languages. We say that a language $L_1 \in \Delta_n^*$ is $\mathbb{C}$-dissectible if there is $L_2 \in \mathbb{C}$ such that L_2 dissects L_1. Let REG denote the family of regular languages. In [12] the REG-dissectibility has been investigated. Several families of REG-dissectible languages have been presented. Moreover it has been shown that there are infinite languages that cannot be dissected with a regular language. Also some open questions for REG-dissectibility can be found in [12]. For example it is not known if the complement of a context free languages is REG-dissectible.

There is a related longstanding open question in [1]: Given two context free languages $L_1, L_2 \subseteq \Delta_n^*$ such that $L_1 \subset L_2$ and $L_2 \setminus L_1$ is an infinite language, is there a context free language L_3 such that $L_3 \subset L_2$, $L_1 \subset L_3$, and both the languages $L_3 \setminus L_1$ and $L_2 \setminus L_3$ are infinite? This question was mentioned also in [12].

Some other results concerning the dissection of infinite languages may be found in [5]. A similar topic is the constructing of minimal covers of languages [2]. Recall that a language $L_1 \subseteq \Delta_n^*$ is called $\mathbb{C}$-*immune* if there is no infinite language $L_2 \subseteq L_1$ such that $L_2 \in \mathbb{C}$. The immunity is also related to the dissection of languages; some results on this theme can be found in

[3, 7, 11].

Let $\mathbb{N}$ denote the set of all positive integers. An infinite language $L \subseteq \Delta_n^*$ is called *constantly growing*, if there is a constant $c_0 \in \mathbb{N}$ and a finite set $K \subset \mathbb{N}$ such that for each $w \in L$ with $|w| \geq c_0$ there is a word $\bar{w} \in L$ and a constant $c \in K$ such that $|\bar{w}| = |w| + c$. We say also that L is (c_0, K)-constantly growing. In [12], it has been proved that every constantly growing language L is REG-dissectible.

We define a *tetration* function (a repeated exponentiation) as follows: $\exp^{1,\alpha} = 2^\alpha$ and $\exp^{j+1,\alpha} = 2^{\exp^{j,\alpha}}$, where $j \in \mathbb{N}$. The tetration function is known as a fast growing function. If k, α are positive positive integers and $L \subseteq \Delta_n^*$ is an infinite language such that for each $u \in L$ there is $v \in L$ with $|u| < |v| \leq \exp^{k,\alpha} |u|$ then we call L a language with the *growth bounded by* (k, α)-tetration.

Let $L \subseteq \Delta_n^*$ be an infinite language with the growth bounded by (k, α)-tetration, where $k \geq 2$. In the current article we show that there are:

- an alphabet Σ_{2k-1} with $|\,\Sigma_{2k-1}\,| = 2k - 1$,
- an erasing alphabetical homomorphism $\upsilon : \Sigma_{2k-1}^* \to \Delta_1^*$,
- a nonerasing alphabetical homomorphism $\pi : \Delta_n^* \to \Delta_1^*$, and
- $3k - 3$ context free languages $L_1, L_2, \ldots, L_{3k-3} \subseteq \Sigma_{2k-1}^*$

such that the homomorphic image $\upsilon(\bigcap_{i=1}^{3k-3} L_i)$ dissects the homomorphic image $\pi(L)$. Thus we may say that the languages with the growth bounded by a (k, α)-tetration are CFL_{3k-3}-dissectible.

We sketch the basic ideas of our proof. Recall that a non-associative word on the letter z is a "well parenthesized" word containing a given number of occurrences of z. It is known that the number of non-associative words containing $n + 1$ occurrences of z is equal to the n-th Catalan number [8]. For example for $n = 3$ we have 5 distinct non-associative words: $(((zz)z)z)$, $((zz)(zz))$, $(z(z(zz)))$, $(z((zz)z))$, and $((z(zz))z)$. Every non-associative word contains the prefix $(^k z$ for some $k \in \mathbb{N}$, where $(^k$ denotes the k-th power of the opening bracket. It is easy to verify that there are non-associative words such that k equals "approximately" $\log_2 n$. We construct three context free languages whose intersection accepts such words; we call these words *balanced non-associative words*. By counting the number of opening brackets of a balanced non-associative word with n occurrences of z we can compute a logarithm of n.

Let $\log_2^{(1)} n = \log_2 n$ and $\log_2^{(j+1)} n = \log_2^{(j)}(\log_2 n)$. Our construction can be "chained" so that we construct $3k-3$ context free languages, whose intersection accepts words with n occurrences of z and a prefix $x^j\bar{z}$, where j is equal "approximately" to $\log_2^{(k)} n$ and $\bar{z} \neq x$. If L is a language with the growth bounded by a (k,α)-tetration then the language $\bar{L} = \{x^j \mid j = \lceil \log_2^{(k)} |w| \rceil \text{ and } w \in L\}$ is constantly growing. Less formally said, by means of intersection of $3k-3$ context free languages we transform the challenge of dissecting a language with the growth bounded by (k,α)-tetration to the challenge of dissecting a constantly growing language. This approach allows us to prove our result.

2 Preliminaries

Let $\mathbb{R}^+$ denote the set of all positive real numbers.

Let $\mathrm{B}_k = \{x_1, x_2, \ldots, x_k\}$ be an ordered alphabet (set) of k distinct opening brackets, and let $\bar{\mathrm{B}}_k = \{y_1, y_2, \ldots, y_k\}$ be an ordered alphabet (set) of k distinct closing brackets. We define the alphabet $\Sigma_{2k-1} = \mathrm{B}_k \cup (\bar{\mathrm{B}}_k \setminus \{y_1\})$. The alphabet Σ_{2k-1} contains all opening brackets B_k and all the closing brackets without the the first one $\bar{\mathrm{B}}_k \setminus \{y_1\}$. It follows that $|\,\Sigma_{2k-1}\,| = 2k-1$.

Let ϵ denote the empty word. Given a finite alphabet S, let S^+ denote the set of all finite nonempty words over the alphabet S and let $S^* = S^+ \cup \{\epsilon\}$.

Let $\mathrm{Fac}(w)$ denote the set of all factors a word $w \in S^*$. We define that $\epsilon, w \in \mathrm{Fac}(w)$; i.e. the empty word and the word w are factors of w. Let $\mathrm{Pref}(w) \subseteq \mathrm{Fac}(w)$ denote the set of all prefixes of $w \in S^*$. We define that $\epsilon, w \in \mathrm{Pref}(w)$. Let $\mathrm{Suf}(w) \subseteq \mathrm{Fac}(w)$ denote the set of all suffixes of $w \in S^*$. We define that $\epsilon, w \in \mathrm{Suf}(w)$. Given a finite alphabet S, let $\mathrm{occur}(w,t)$ denote the number of occurrences of the nonempty factor $t \in S^+$ in the word $w \in S^*$; formally $\mathrm{occur}(w,t) = |\{v \in \mathrm{Suf}(w) \mid t \in \mathrm{Pref}(v)\}|$.

Given two finite alphabets S_1, S_2, a *homomorphism* from S_1^* to S_2^* is a function $\tau : S_1^* \to S_2^*$ such $\tau(ab) = \tau(a)\tau(b)$, where $a, b \in S_1^+$. It follows that in order to define a homomorphism τ, it suffices to define $\tau(z)$ for every $z \in S_1$; such definition "naturally" extends to every word $a \in S_1^+$. We say that τ is a *nonerasing alphabetical homomorphism* if $\tau(z) \in S_2$ for every $z \in S_1$. We say that τ is an *erasing alphabetical homomorphism* if $\tau(z) \in S_2 \cup \{\epsilon\}$ for every $z \in S_1$ and there is at least one $z \in S_1$ such that $\tau(z) = \epsilon$.

3 Balanced non-associative words

Suppose $k, m \in \mathbb{N}$, where $k, m \geq 2$, and $k \geq m$. To simplify the notation we define $x = x_m$, $y = y_m$, and $z = x_{m-1}$; it means that x denotes the m-th opening bracket, y denotes the m-th closing bracket, and z denotes the $m-1$-th opening bracket.

Let $\mu_{k,m} : \Sigma_{2k-1}^* \rightarrow \Sigma_{2k-1}^*$ be an erasing alphabetical homomorphism defined as follows:

- $\mu_{k,m}(z) = z$,
- $\mu_{k,m}(x) = x$,
- $\mu_{k,m}(y) = y$.
- $\mu_{k,m}(a) = \epsilon$, where $a \in \Sigma_{2k-1} \setminus \{x, y, z\}$.

Given a language $L \subseteq \Sigma_{2k-1}^*$, we define the language $\mu_{k,m}(L) = \{\mu_{k,m}(w) \mid w \in L\}$.

Remark 3.1. *For given k, m the erasing alphabetical homomorphism $\mu_{k,m}$ sends all opening and closing brackets from B_k and $\bar{\mathrm{B}}_k$ to the empty string with the exception of x, y, and z.*

Let $\mathrm{Naw}_{k,m} \subseteq \Sigma_{2k-1}^*$ be the context free language generated by the following context free grammar, where S is a start non-terminal symbol, N is a non-terminal symbol, and x, y, z, a are terminal symbols (the letters from Σ_{2k-1}).

- $\mathrm{S} \rightarrow \mathrm{N}\, x\, \mathrm{N}\,\mathrm{S}\,\mathrm{S}\,\mathrm{N}\, y\, \mathrm{N} \mid \mathrm{N}\, x\, \mathrm{N}\, z\, \mathrm{N}\, y\, \mathrm{N} \mid \mathrm{N}\, x\, \mathrm{N}\, z\, \mathrm{N}\, z\, \mathrm{N}\, y\, \mathrm{N}$,
- $\mathrm{N} \rightarrow a\, \mathrm{N} \mid \epsilon$, where $a \in \Sigma_{2k-1} \setminus \{x, y, z\}$.

We call the words from $\mathrm{Naw}_{k,m}$ *non-associative words* over the opening bracket x, the closing bracket y, and the letter z.

Remark 3.2. *Let $M = \mu_{k,m}(\mathrm{Naw}_{k,m})$. To understand the definition of $\mathrm{Naw}_{k,m}$, note that the language M is generated by the context free grammar defined by: $\mathrm{S} \rightarrow x\,\mathrm{S}\,\mathrm{S}\,y \mid xzy \mid xzzy$. To see this, just remove the non-terminal symbol N in the definition of $\mathrm{Naw}_{k,m}$. The usage of the non-terminal symbol N allows to "insert" between any two letters of a word from $\mu_{k,m}(\mathrm{Naw}_{k,m})$ the words from $K = (\Sigma_{2k-1} \setminus \{x, y, z\})^*$; the set K contains* *words from* Σ_{2k-1}^* *that have no occurrence of* x, y, z. *It means that if* $w = w_1 w_2 \dots w_n \in \mu_{k,m}(\mathrm{Naw}_{k,m})$, *then* $t_0 w_1 t_1 w_2 t_2 \dots t_{n-1} w_n t_n \in \mathrm{Naw}_{k,m}$, *where* $w_i \in \{x, y, z\}$ *and* $t_i \in K$.

The reason for the name "non-associative words" is the obvious similarity between the words from M *and the "standard non-associative words" mentioned in the introduction section. Our definition guarantees that* $w_1 xzyw_2 \in M$ *if and only if* $w_1 xzzyw_2 \in M$ *for every* $w_1, w_2 \in \{x, z, y\}^*$.

Recall that a pushdown automaton is a 6-tuple $(\mathrm{Q}, \Delta, \Gamma, q_0, \mathrm{S}, \delta)$, where

- Q is a set of states,
- Δ is an input alphabet,
- Γ is a stack alphabet,
- $q_0 \in \mathrm{Q}$ is an input state,
- $\mathrm{S} \in \Gamma$ is the initial symbol of the stack,
- $\delta : (\mathrm{Q} \times \Delta \times \Gamma) \to (\mathrm{Q}, \Gamma^*)$ is a transition function.

We define that a pushdown automaton accepts a word by the empty stack, hence we do not need to define the set of final states. Given a pushdown automaton g, let $\mathrm{AL}(g) \subseteq \Delta^*$ denotes the language accepted by g.

Let $\Lambda_{k,m} = \mathrm{AL}(g_{k,m}) \subseteq \Sigma_{2k-1}^*$ denote the context free language accepted by the pushdown automaton $g_{k,m} = (\mathrm{Q}, \Sigma_{2k-1}, \Gamma, q_S, \mathrm{S}, \delta)$, where:

- $\mathrm{Q} = \{q_S, q_B, q_0, q_x, q_r\}$,
- $\Gamma = \{\mathrm{S}, X\}$,
- $\delta(q, a, u) \to (q, u)$, where $q \in \mathrm{Q}$, $u \in \Gamma$, and $a \in \Sigma_{2k-1} \setminus \{x, y, z\}$,
- $\delta(q_S, x, u) \to (q_B, u)$, where $u \in \Gamma$,
- $\delta(q_S, z, u) \to (q_S, u)$, where $u \in \Gamma$,
- $\delta(q_S, y, u) \to (q_S, u)$, where $u \in \Gamma$,
- $\delta(q_B, x, u) \to (q_x, uXX)$, where $u \in \Gamma$,
- $\delta(q_B, z, u) \to (q_S, u)$, where $u \in \Gamma$,

- $\delta(q_B, y, u) \to (q_S, u)$, where $u \in \Gamma$,
- $\delta(q_0, x, u) \to (q_x, u)$, where $u \in \Gamma$,
- $\delta(q_0, z, u) \to (q_0, u)$, where $u \in \Gamma$,
- $\delta(q_0, y, u) \to (q_0, u)$, where $u \in \Gamma$,
- $\delta(q_x, x, u) \to (q_x, uX)$, where $u \in \Gamma$,
- $\delta(q_x, z, X) \to (q_0, \epsilon)$,
- $\delta(q_x, z, \mathrm{S}) \to (q_r, X)$, where $u \in \Gamma$,
- $\delta(q_x, y, u) \to (q_0, u)$, where $u \in \Gamma$, and
- $\delta(q_r, a, u) \to (q_r, u)$, where $r \in \Sigma_{2k-1}$ and $u \in \Gamma$.

Remark 3.3. *Note in the definition of $g_{k,m}$ that the letters from $\Sigma_{2k-1} \setminus \{x, y, z\}$ change neither the state of $g_{k,m}$ nor the stack. Hence to illuminate the behavior of $g_{k,m}$, we can consider only words over the alphabet $\{x, y, z\}$. Then it is easy to see that the pushdown automaton $g_{k,m}$ pushes XX on the stack on the first occurrence of xx. For every other occurrence of xx the pushdown automaton $g_{k,m}$ pushes X on the stack. Once reached the state q_x, then for every occurrence of xz one X is removed from the stack. The state q_r works as a refuse state. Note that after reaching the state q_r the stack is not empty, the stack cannot be changed, and no other state can be reached from q_r. The states q_S and q_B enable to recognize the first occurrence of xx. Once the states q_x are reached, the states q_S and q_B can not be reached any more.*

Thus the pushdown automaton $g_{k,m}$ accepts all words, where the number of occurrences of xz after the first occurrence of xx is exactly one more than the number of occurrences of xx. Formally, if $w \in \mu_{k,m}(\Sigma_{2k-1}^*)$ then we define $\bar{w}$ as follows:

- *If* $\mathrm{occur}(w, xx) = 0$ *then* $\bar{w} = \epsilon$.
- *If* $\mathrm{occur}(w, xx) \geq 1$ *then let* $\bar{w} \in \mathrm{Suf}(w)$ *be such that* $xx \in \mathrm{Pref}(\bar{w})$ *and* $\mathrm{occur}(\bar{w}, xx) = \mathrm{occur}(w, xx)$.

Clearly $\bar{w}$ is uniquely defined. Then we have that $w \in \mu_{k,m}(\Lambda_{k,m})$ if and only if $\bar{w} = \epsilon$ or $\mathrm{occur}(\bar{w}, xx) + 1 = \mathrm{occur}(\bar{w}, xz)$. It follows that the words without any occurrence of xx are accepted. In the following we will consider *the words from the intersection $U = \Lambda_{k,m} \cap \mathrm{Naw}_{k,m}$. Note that there are only two nonempty words $xzy, xzzy \in U$, that have no occurrence of xx.*

Recall that a "standard" non-associative word can be represented as a full binary rooted tree graph, where every inner node represents a corresponding pair of brackets and every leaf represents the letter z [8]. It is known that the number of inner nodes plus one is equal to the number of leaves in a full binary rooted tree graph. In the case of non-associative words from $\mathrm{Naw}_{k,m}$, let the leaves represent the factors xzy and $xzzy$. Then the number of occurrences of xz is equal to the number of leaves and the number of occurrences of xx is equal to the number of inner nodes. Hence the intersection $M \cap \mathrm{Naw}_{k,m}$ contains non-associative words that have no "unnecessary" brackets; for example $xzzy, xxzzzyy, xxxzzzyyy \in \mathrm{Naw}_{k,m}$, $xzzy \in M$ and $xxzzzyy, xxxzzzyyy \notin M$.

Let $\mathrm{Bal}_{k,m} \subseteq \Sigma_{2k-1}^*$ be the context free language generated by the following context free grammar, where S is a start non-terminal symbol, N, K, V, P are non-terminal symbols, and x, y, z, a are terminal symbols (the letters from Σ_{2k-1}).

- $\mathrm{S} \rightarrow KVP$,
- $V \rightarrow VV \mid \mathrm{N}\,z\,\mathrm{N} \mid \mathrm{N}\,zTz\,\mathrm{N} \mid \epsilon$,
- $T \rightarrow \mathrm{N}\,y\,\mathrm{N}\,T\,\mathrm{N}\,x\,\mathrm{N} \mid \epsilon$,
- $K \rightarrow KK \mid \mathrm{N}\,x\,\mathrm{N} \mid \epsilon$,
- $P \rightarrow PP \mid \mathrm{N}\,y\,\mathrm{N} \mid \epsilon$,
- $\mathrm{N} \rightarrow a\,\mathrm{N} \mid \epsilon$, where $a \in \Sigma_{2k-1} \setminus \{x, y, z\}$.

We call the words from $\mathrm{Bal}_{k,m}$ *balanced words*.

Remark 3.4. *Let $M = \mu_{k,m}(\mathrm{Bal}_{k,m})$. It is easy to see that the words from the language M contains no factor of the form zy^ix^jz, where i, j are distinct positive integers; hence the name "balanced" words. The non-terminal symbols K, P enable that if $w \in M$ then w has a prefix x^i and a suffix y^j for all $i, j \in \mathbb{N} \cup \{0\}$.*

The non-terminal symbol N in the definition of $\mathrm{Bal}_{k,m}$ has the same purpose like in the definition of $\mathrm{Naw}_{k,m}$.

Let

$$\Omega_{k,m} = \mathrm{Naw}_{k,m} \cap \mathrm{Bal}_{k,m} \cap \Lambda_{k,m}.$$

We call the words from $\Omega_{k,m}$ *balanced non-associative words* over the opening bracket x, the closing bracket y, and a letter z.

Let $\Omega_{k,m}(n) = \{w \in \Omega_{k,m} \mid \mathrm{occur}(w,z) = n\}$, where $n \in \mathbb{N}$. The set $\Omega_{k,m}(n)$ contains the balanced non-associative words having exactly n occurrences of the letter z.

Given a word $w \in \Sigma_{2k-1}^*$ and $a \in \Sigma_{2k-1}$, let

$$\mathrm{height}(w,a) = \max\{j \mid a^j \in \mathrm{Fac}(w)\}.$$

The height of a word w is the maximal power of the letter a, that is a factor of w. We show that if $w \in \mu_{k,m}(\Omega_{k,m})$ and h is the height the opening bracket x in w then x^h is a prefix of w and y^h is a suffix of w.

Lemma 3.5. *If $w \in \mu_{k,m}(\Omega_{k,m})$ and $h = \mathrm{height}(w,x)$ then $x^h \in \mathrm{Pref}(w)$ and $y^h \in \mathrm{Suf}(w)$.*

Proof. Note that $\mu_{k,m}(\Omega_{k,m}) \subseteq \Omega_{k,m}$. Since $\Omega_{k,m} \subseteq \mathrm{Naw}_{k,m}$, there is $\bar{h} \in \mathbb{N}$ such that $x^{\bar{h}}z \in \mathrm{Pref}(w)$. To get a contradiction suppose that $\bar{h} < h$. Because $\Omega_{k,m} \subseteq \mathrm{Bal}_{k,m}$ it follows that $w = x^{\bar{h}}w_1zy^hx^hzw_2$ for some $w_1 \in \mathrm{Fac}(w)$ with $z \in \mathrm{Pref}(w_1z)$ and $w_2 \in \mathrm{Suf}(w)$.

Consider the prefix $r = x^{\bar{h}}w_1zy^h$. Obviously $w_1z \in \mu_{k,m}(\mathrm{Bal}_{k,m})$. It is easy to see that if $v \in \mu_{k,m}(\mathrm{Bal}_{k,m})$, $x \notin \mathrm{Pref}(v)$, and $y \notin \mathrm{Suf}(v)$ then $\mathrm{occur}(v,x) = \mathrm{occur}(v,y)$. Thus $\mathrm{occur}(w_1z,x) = \mathrm{occur}(w_1z,y)$. It follows that $\mathrm{occur}(r,x) < \mathrm{occur}(r,y)$.

This is a contradiction, since for every prefix $v \in \mathrm{Pref}(w)$ of a non-associative word $w \in \mathrm{Naw}_{k,m}$ we have that $\mathrm{occur}(v,x) \geq \mathrm{occur}(v,y)$. We conclude that $\bar{h} = h$ and $x^h \in \mathrm{Pref}(w)$. In an analog way we can show that $y^h \in \mathrm{Suf}(w)$. This completes the proof. □

For a word $w \in \mu_{k,m}(\Omega_{k,m})$, we show the relation between the height of w and the number of occurrences of z in w.

Proposition 3.6. *If $w \in \mu_{k,m}(\Omega_{k,m})$ and $h = \mathrm{height}(w,x)$ then*

$$2^{h-1} \leq \mathrm{occur}(w,z) \leq 2^h.$$

Proof. We prove the proposition for all h by induction:

- If $h = 0$ then $w = \epsilon$.
- If $h = 1$ then $w \in \{xzzy, xzy\}$.
- If $h = 2$ then $w \in \{xxzyxzyy, xxzzyxzyy, xxzyxzzyy, xxzzyxzzyy\}$.

Thus the proposition holds for $h \leq 2$. Since $\Omega_{k,m} \subseteq \Lambda_{k,m}$, clearly we have that if $h \geq 2$ then $w = xw_1w_2y$, where $w_1, w_2 \in \mu_{k,m}(\Omega_{k,m})$. Suppose the proposition holds for all $\bar{h} < h$. We prove the proposition holds for h.

Let $h_1 = \text{height}(w_1, x)$ and $h_2 = \text{height}(w_2, x)$. Lemma 3.5 implies that $x^{h_1} \in \text{Pref}(w_1)$, $y^{h_1} \in \text{Suf}(w_1)$, $x^{h_2} \in \text{Pref}(w_2)$, and $y^{h_2} \in \text{Suf}(w_2)$. Since $w \in \mu_{k,m}(\text{Bal}_{k,m})$ it follows that $h_1 = h_2$. Because $x^{h_1} \in \text{Pref}(w_1)$ we have that $x^{h_1+1} \in \text{Pref}(w)$. Clearly $\text{occur}(w, x^{h_1+1}) = 1$; note that $\text{occur}(w_1w_2, x^{h_1+1}) = 0$. Thus $h_1 + 1 = h$. For we assumed that the proposition holds for all $\bar{h} < h$, we can derive that

$$\text{occur}(w, z) = \text{occur}(w_1, z) + \text{occur}(w_2, z) \leq 2^{h_1} + 2^{h_1} = 2^{h_1+1} = 2^h$$

and

$$\text{occur}(w, z) = \text{occur}(w_1, z) + \text{occur}(w_2, z) \geq 2^{h_1-1} + 2^{h_1-1} = 2^{h_1} = 2^{h-1}.$$

This completes the proof. □

Proposition 3.6 have the following obvious corollary.

Corollary 3.7. *If $n \in \mathbb{N}$, $w \in \mu_{k,m}(\Omega_{k,m}(n))$, and $h = \text{height}(w, x)$ then*

$$\log_2 n \leq h \leq 1 + \log_2 n.$$

Given $w, u, v \in \Sigma_{2k-1}^{+}$, let $\text{replace}(w, v, u)$ denote the word built from w by replacing the first occurrence of v in w by u. Formally, if $\text{occur}(w, v) = 0$ then $\text{replace}(w, v, u) = w$. If $\text{occur}(w, v) = j > 0$ and $w = w_1vw_2$, where $\text{occur}(vw_2, v) = j$ then $\text{replace}(w, v, u) = w_1uw_2$.

We prove that the set of balanced non-associative words $\Omega_{k,m}(n)$ having n occurrences of z is nonempty for each $n \in \mathbb{N}$.

Lemma 3.8. *If $n \in \mathbb{N}$ then $\Omega_{k,m}(n) \neq \emptyset$.*

Proof. If $n = 1$ then $xzy \in \Omega_{k,m}(1)$. Given $n \in \mathbb{N}$ with $n > 1$, let $j \in \mathbb{N}$ be such that $2^{j-1} < n \leq 2^j$. Obviously such j exists and is uniquely determined. Let $w_1 = xzzy$. Let $w_{i+1} = xw_iw_iy$. Clearly $\operatorname{occur}(w_j, z) = 2^j$ and $w_j \in \Omega_{k,m}(2^j)$. Note that $\operatorname{occur}(w_j, xzzy) = 2^{j-1}$. Let $w_{j,0} = w_j$ and $w_{j,i+1} = \operatorname{replace}(w_{j,i}, xzzy, xzy)$, where $i \in \mathbb{N} \cup \{0\}$ and $i \leq 2^{j-1}$. Let $\alpha = 2^j - n$. Then one can easily verify that $\operatorname{occur}(w_{j,\alpha}, z) = n$ and $w_{j,\alpha} \in \Omega_{k,m}(n)$.

Less formally said, we construct a balanced non-associative word w_j having 2^{j-1} occurrences of $xzzy$ and then we replace a given number of occurrences of $xzzy$ with the factor xzy to achieve the required number of occurrences of z. This completes the proof. □ □

4 Intersection of balanced non-associative words

Let $\Omega_k = \bigcap_{m=2}^{k} \Omega_{k,m}$ and let $\Omega_k(n) = \{w \in \Omega_k \mid \operatorname{occur}(w, x_1) = n\}$. We show that for all positive integers n, k with $k \geq 2$ there is a word $w \in \Omega_k$ such that w has n occurrences of the opening bracket x_1.

Proposition 4.1. *If $k, n \in \mathbb{N}$ and $k \geq 2$ then $\Omega_k(n) \neq \emptyset$.*

Proof. Let $h(1) = n$. Let $w_i \in \mu_{k,i}(\Omega_{k,i}(h(i-1))$ and let $h(i) = \operatorname{height}(w_i, x_i)$, where $i \in \{2, 3, 4, \dots, k\}$. Lemma 3.8 implies that such w_i exist.

Let $v_2 = w_2$. Let $v_{j+1} = \operatorname{replace}(v_j, x_j^{h(j)}, w_{j+1})$, where $j \in \mathbb{N}$ and $j \geq 2$. Lemma 3.5 implies that $x^{h(j)} \in \operatorname{Pref}(v_j)$. Note that $\mu_{k,j}(v_j + 1) = \mu_{k,j}(v_j)$. Then it is quite straightforward to see that $v_k \in \Omega_k$ and $\operatorname{occur}(v_k, x_1) = n$. Less formally said, with every iteration we construct a non-associative word by "well parenthesizing" the prefix $x_i^{h(i)}$ with the opening bracket x_{i+1} and the closing bracket y_{i+1}. This completes the proof. □ □

To clarify the proof of Proposition 4.1, let us see the following example.

Example 4.2. *Let $n = 23$ and $k = 4$. To make the example easy to read, we define* $\mathrm{B}_4 = \{z, (, [, <\}$ *and* $\bar{\mathrm{B}}_4 = \{\bar{z},),], >\}$. *It means that* $x_1 = z$, $x_2 = ($, $x_3 = [$, $x_4 =<$, $\bar{x}_1 = \bar{z}$, $\bar{x}_2 =)$, $\bar{x}_3 =]$, *and* $\bar{x}_4 =>$.

To fit the example into the width of the page, we define auxiliary words u_1 and u_2:

- $u_1 = z)(z))((z)(z)))(((z)(z))((z)(z))))$,
- $u_2 = ((((z)(zz))((zz)(zz)))(((zz)(zz))((zz)(zz)))))$.

Then we have that

- $h(1) = 23$; $w_2 = (((((u_1u_2$; $h(2) = 5$; $w_3 = [[[(][(]][[(][((]]$;
- $h(3) = 3$; $w_4 =<< [[>< [[>>$; $h(4) = 2$; $v_2 = w_2$; $v_3 = [[[(][(]][[(][((]]u_1u_2$;
- $v_4 =<< [>< [[>> (][(]][[(][((]]]u_1u_2$.

This ends the example.

We define two technical functions $\log_2^{(j)} t$ and $\log_2^{[j]} t$ for all $j \in \mathbb{N}$ and $t \in \mathbb{R}^+$ as follows:

- $\log_2^{(1)} t = \log_2 t$ and $\log_2^{(j+1)} t = \log_2^{(j)} (\log_2 t)$.
- $\log_2^{[1]} t = 1 + \log_2 t$ and $\log_2^{[j+1]} t = \log_2^{[j]} (1 + \log_2 t)$.

It is a simple exercise to prove the following lemma. We omit the proof.

Lemma 4.3. *If $j \in \mathbb{N}$ then for each $t \in \mathbb{R}^+$ with $t \geq 1$ we have that*

$$\log_2^{(j)} t \leq \log_2^{[j]} t \leq j + \log_2^{(j)} t.$$

Using the function $\log_2^{(k)} t$ we present an upper and a lower bound for the height of words from Ω_k.

Proposition 4.4. *If $k \in \mathbb{N}$ and $k \geq 2$ then for each $w \in \Omega_k$, $h = \mathrm{height}(w, x_k)$, and $n = \mathrm{occur}(w, x_1)$ we have*

$$\log_2^{(k)} n \leq h \leq k + \log_2^{(k)} n.$$

Proof. It follows from Corollary 3.7 that $\log_2^{(k)} n \leq h \leq \log_2^{[k]} n$. Then the proposition follows from Lemma 4.3. $\square$ $\square$

5 Dissection by a regular language

In [12] it was shown that every constantly growing language can be dissected by some regular language.

Lemma 5.1. *(see [12, Lemma 3.3]) Every infinite constantly growing language is* REG*-dissectible.*

From the proof of Lemma 3.3 in [12] we can formulate the following Lemma.

Lemma 5.2. *If $n, c_0 \in \mathbb{N}$, $K \subset \mathbb{N}$, $|K| < \infty$, $c = \max\{j \in K\}$, and $L \subseteq \Delta_n^*$ is a (c_0, K)-constantly growing language then there are $j_1, j_2 \in \{0, 1, 2, \ldots, c\}$ such that $j_1 \neq j_2$ and both sets H_1, H_2 are infinite, where*

$$H_i = \{w \mid w \in L \text{ and } |w| \equiv j_i \pmod{c+1}\} \text{ and } i \in \{1, 2\}.$$

6 Tetration

Recall that a deterministic finite automaton g is 5-tuple $(\mathrm{Q}, \Delta, q_0, \delta, \mathrm{F})$, where Q is the set of states, Δ is an input alphabet, q_0 is the initial state, δ is a transition function, and F is the set of accepting states. Let $\mathrm{AL}(g)$ denote the language accepted by g; $\mathrm{AL}(g)$ is a regular language.

We prove that if $L \subseteq \Omega_k$ is an infinite language of balanced non-associative words with the number of occurrences of x_1 "bounded" by (k, α)-tetration then L can be dissected by a regular language.

Proposition 6.1. *If $k, \alpha \in \mathbb{N}$, $k \geq 2$, and $L \subseteq \Omega_k$ is an infinite language such that for each $w_1 \in L$ there is $w_2 \in L$ with $\mathrm{occur}(w_1, x_1) < \mathrm{occur}(w_2, x_2)$ and $\mathrm{occur}(w_2, x_1) \leq \exp^{k,\alpha} \mathrm{occur}(w_1, x_1)$ then there is a regular language R such that R dissects L and the minimal deterministic finite automaton accepting R has at most $k + \alpha + 3$ states.*

Proof. Let $w_1, w_2 \in L$ be such that

$$n_2 \leq \exp^{k,\alpha} n_1, \tag{1}$$

where $n_1 = \mathrm{occur}(w_1, x_1)$ and $n_2 = \mathrm{occur}(w_2, x_1)$.

Let $h_1 = \mathrm{height}(\mu_{k,k}(w_1), x_k)$ and $h_2 = \mathrm{height}(\mu_{k,k}(w_2), x_k)$. Proposition 4.4 implies that

$$\log_2^{(k)} n_1 \leq h_1 \text{ and } h_2 \leq k + \log_2^{(k)} n_2 \tag{2}$$

From (1) and (2) we have that

$$h_2 \leq k + \log_2^{(k)} n_2 \leq k + \log_2^{(k)}(\exp^{k,\alpha} n_1). \tag{3}$$

Realize that $\log_2(\exp^{j,\alpha}) = \exp^{j-1,\alpha}$ and that if $a, b \in \mathbb{R}^+$ and $a, b \geq 2$ then $a + b \leq ab$. Then we have that

$$\log_2^{(j)}(\exp^{j,\alpha} n_1) = \log_2^{(j-1)}(\exp^{j-1,\alpha} + \log_2 n_1) \leq \log_2^{(j-1)}(\exp^{j-1,\alpha} \log_2 n_1). \tag{4}$$

From (4) it follows that

$$\log_2^{(k)}(\exp^{k,\alpha} n_1) \leq \log_2(\exp^{1,\alpha} \log_2^{(k-1)} n_1) = \alpha + \log_2^{(k)} n_1. \tag{5}$$

From (2), (3), and (5) we have that

$$h_2 \leq k + \alpha + \log_2^{(k)} n_1 \leq k + \alpha + h_1. \tag{6}$$

The equation (6) says that for each $u \in L$ there is $v \in L$ with $|u| < |v|$ and $\mathrm{height}(\mu_{k,k}(v), x_k) \leq k + \alpha + \mathrm{height}(\mu_{k,k}(u), x_k)$.

Lemma 5.2 implies that there are distinct non-negative integers $j_1, j_2 \leq k + \alpha$ such that both H_1, H_2 are infinite sets, where

$$H_i = \{v \mid v \in L \text{ and } \mathrm{height}(\mu_{k,k}(v), x_k) \equiv j_i \pmod{k+\alpha+1}\} \text{ and } i \in \{1, 2\}.$$

Let $c = k+\alpha$. Consider the deterministic finite automaton $g = (\mathrm{Q}, \Sigma_{2k-1}, q_0, \delta, \mathrm{F})$, where

- $\mathrm{Q} = \{q_0, q_1, \ldots, q_c, q_a, q_r\}$,
- $\delta(q, x) \to (q)$, where $q \in \mathrm{Q}$ and $x \in \Sigma_{2k-1} \setminus \{x_k, x_{k-1}\}$,
- $\delta(q_i, x_k) \to (q_{i+1 \mod c+1})$,
- $\delta(q_{j_1}, x_{k-1}) \to (q_a)$,
- $\delta(q_i, x_{k-1}) \to (q_r)$, where $i \neq j_1$,
- $\delta(q, x) \to (q)$, where $q \in \{q_a, q_r\}$ and $x \in \{x_k, x_{k-1}\}$, and
- $\mathrm{F} = \{q_{j_1}, q_a\}$.

The deterministic finite automaton g implements the modulo operation on the prefix of the form x_k^i. The input letter $x \in \Sigma_{2k-1} \setminus \{x_k, x_{k-1}\}$ does not change the state. The input letter x_k changes the state from q_i to $q_{i+1 \mod c+1}$. If the input letter equals x_{k-1} then the state changes either to accept q_a or refuse q_r. Realize that if $w \in \mu_{k,k}(\Omega_k)$, $a \in \{y_k, x_{k-1}\}$, and $x_k a \in \mathrm{Fac}(w)$

then $a = x_{k-1}$, hence we do not need any "special" transition rule for the letter y_k. Once in the state q_a or q_r, no other states can be reached. The states q_{j_1} and q_a are the accepting states. It is easy to see that $\mathrm{AL}(g) = H_1$ and in consequence the regular language $R = \mathrm{AL}(g)$ dissects L.

This completes the proof. $\square$ $\square$

Given $n \in \mathbb{N}$, let Δ_n be some alphabet with n letters. Let $\Delta_1 = \mathrm{B}_1 = \{x_1\}$ be the alphabet with the "first" opening bracket. Let $L \subseteq \Delta_n^*$ be an infinite language with a growing bounded by (k,α)-tetration. Let $\upsilon : \Sigma_{2k-1}^* \rightarrow \Delta_1$ be an erasing alphabetical homomorphism defined by $\upsilon(x_1) = x_1$ and $\upsilon(a) = \epsilon$, where $a \in \Sigma_{2k-1} \setminus \{x_1\}$. Let $\pi : \Delta_n^* \rightarrow \Delta_1$ be a nonerasing alphabetical homomorphism defined by $\pi(a) = x_1$ for all $a \in \Delta_n$. Note that if $w \in \Delta_n^*$ then $|w| = |\pi(w)|$.

We show that there $3k-3$ context free languages $L_1, L_2, \dots, L_{3k-3} \subseteq \Sigma_{2k-1}^*$ such that the homomorphic image $\upsilon(\bigcap_i^{3k} L_i)$ dissects the homomorphic image $\pi(L)$.

Theorem 6.2. *If $n, \alpha, k \in \mathbb{N}$, $k \geq 2$, $L \subseteq \Delta_n^*$ is an infinite language with the growth bounded by (k,α)-tetration then there are $3k-3$ context free languages $L_1, L_2, \dots, L_{3k-3}$ such that $\upsilon(\bigcap_i^{3k-3} L_i)$ dissects $\pi(L)$.*

Proof. Recall that the language Ω_k is an intersection of $3k-3$ context free languages:

$$\Omega_k = \bigcap_{m=2}^{k} \left(\mathrm{Naw}_{k,m} \cap \mathrm{Bal}_{k,m} \cap \Lambda_{k,m}\right).$$

Let us denote these languages $\tilde{L}_1, \tilde{L}_2, \dots, \tilde{L}_{3k-3}$.

Let $\pi(L) = \{\pi(w) \mid w \in L\} \subseteq \Delta_1^*$ and let $\bar{L} = \{w \in \Omega_k \mid \upsilon(w) \in \pi(L)\} \subseteq \Omega_k$. Note that $\bar{L}$ contains $w \in \Omega_k$ if and only if there is $\bar{w} \in L$ such that the number of occurrences of x_1 in w is equal to the length of $\bar{w}$; formally $\mathrm{occur}(w, x_1) = |\bar{w}|$.

Since L is a language with the growth bounded by (k,α)-tetration, we have that for each $w_1 \in \bar{L}$ there is $w_2 \in \bar{L}$ with $\mathrm{occur}(w_2, x_1) \leq \exp^{k,\alpha} \mathrm{occur}(w_1, x_1)$. Then Proposition 6.1 implies that there is a regular language R that dissects $\bar{L}$. It is well known that intersection of a regular language and a context free language is a context free language. Hence let $L_1 = \tilde{L}_1 \cap R$ and let $L_j = \tilde{L}_j$ for all $j \geq 2$ and $j \leq 3k-3$. Then $\bigcap_{i=1}^{3k-3} L_i$ dissects $\bar{L}$. The theorem follows.
$\square$ $\square$

Acknowledgments

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS20/183/OHK4/3T/14.

References

[1] W. Bucher, *A density problem for context-free languages*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS 10, (1980).

[2] M. Domaratzki, J. Shallit, and S. Yu, *Minimal covers of formal languages*, in Developments in Language Theory, 2001.

[3] P. Flajolet and J. M. Steyaert, *On sets having only hard subsets*, in Automata, Languages and Programming, J. Loeckx, ed., Berlin, Heidelberg, 1974, Springer Berlin Heidelberg, pp. 446–457.

[4] S. Ginsburg and S. Greibach, *Deterministic context free languages*, Information and Control, 9 (1966), pp. 620 – 648.

[5] J. Julie, J. Baskar Babujee, and V. Masilamani, *Dissecting power of certain matrix languages*, in Theoretical Computer Science and Discrete Mathematics, S. Arumugam, J. Bagga, L. W. Beineke, and B. Panda, eds., Cham, 2017, Springer International Publishing, pp. 98–105.

[6] L. Liu and P. Weiner, *An infinite hierarchy of intersections of context-free languages*, Math. Systems Theory 7, 185–192., (1973).

[7] E. L. Post, *Recursively enumerable sets of positive integers and their decision problems*, Bull. Amer. Math. Soc., 50 (1944), pp. 284–316.

[8] R. P. Stanley and S. Fomin, *Enumerative Combinatorics*, vol. 2 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1999.

[9] D. Wotschke, *The Boolean Closures of the Deterministic and Nondeterministic Context-Free Languages*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1973, pp. 113–121.

[10] D. Wotschke, *Nondeterminism and boolean operations in pda's*, Journal of Computer and System Sciences, 16 (1978), pp. 456 – 461.

[11] T. Yamakami, *Intersection and union hierarchies of deterministic context-free languages and pumping lemmas*, in Language and Automata Theory and Applications, A. Leporati, C. Martín-Vide, D. Shapira, and C. Zandron, eds., Cham, 2020, Springer International Publishing, pp. 341–353.

[12] T. Yamakami and Y. Kato, *The dissecting power of regular languages*, Information Processing Letters, 113 (2013), pp. 116 – 122.