Palindromic factorization of rich words

Josef Rukavicka

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

December 2022

Finite and Infinite Words, Palindromes

$\mathcal{A}$ - *alphabet* with q letters, where q is finite.

Finite and Infinite Words, Palindromes

$\mathcal{A}$ - *alphabet* with q letters, where q is finite.

A *finite word* of length n is a sequence $u_1 u_2 \cdots u_n$ with $u_i \in \mathcal{A}$. An *infinite word* is a sequence $u_1 u_2 \cdots$ with $u_i \in A$. A *word* means both finite and infinite word. The *empty word* is denoted by ϵ.

Finite and Infinite Words, Palindromes

$\mathcal{A}$ - *alphabet* with q letters, where q is finite.

A *finite word* of length n is a sequence $u_1u_2\cdots u_n$ with $u_i \in \mathcal{A}$. An *infinite word* is a sequence $u_1u_2\cdots$ with $u_i \in A$. A *word* means both finite and infinite word. The *empty word* is denoted by ϵ.

Suppose $u \in \mathcal{A}^n$, $u = u_1u_2\dots u_n$, where $u_i \in \mathcal{A}$.
We define the *reversal* $u^R = u_nu_{n-1}\cdots u_1$.
If $u = u^R$ then u is called a *palindrome*.
We define that ϵ is a palindrome.

Finite and Infinite Words, Palindromes

$\mathcal{A}$ - *alphabet* with q letters, where q is finite.

A *finite word* of length n is a sequence $u_1 u_2 \cdots u_n$ with $u_i \in \mathcal{A}$. An *infinite word* is a sequence $u_1 u_2 \cdots$ with $u_i \in A$. A *word* means both finite and infinite word. The *empty word* is denoted by ϵ.

Suppose $u \in \mathcal{A}^n$, $u = u_1 u_2 \ldots u_n$, where $u_i \in \mathcal{A}$.
We define the *reversal* $u^R = u_n u_{n-1} \cdots u_1$.
If $u = u^R$ then u is called a *palindrome*.
We define that ϵ is a palindrome.

Example

Examples of palindromes: level, noon.

Rich words

Given words u, v, we say that v is a *factor* of u if there are words p, s such that $u = pvs$.

Rich words

Given words u, v, we say that v is a *factor* of u if there are words p, s such that $u = pvs$.

A *palindromic factor* is a factor that is also a palindrome.

Rich words

Given words u, v, we say that v is a *factor* of u if there are words p, s such that $u = pvs$.

A *palindromic factor* is a factor that is also a palindrome.

A word u of length n is called *rich* if u has $n + 1$ distinct palindromic factors. An infinite word w is rich if every finite factor of w is rich.

Rich words

Given words u, v, we say that v is a *factor* of u if there are words p, s such that $u = pvs$.

A *palindromic factor* is a factor that is also a palindrome.

A word u of length n is called *rich* if u has $n + 1$ distinct palindromic factors. An infinite word w is rich if every finite factor of w is rich.

Example

$w = 1001101$
Palindromic factors of w: ϵ (empty word), 1, 0, 00, 1001, 11, 0110, 101.

Properties Rich words I

A finite word w of length n contains at most $n+1$ distinct palindromic factors.

Properties Rich words I

A finite word w of length n contains at most $n+1$ distinct palindromic factors.

Every factor of a rich word is also rich.

Properties Rich words I

A finite word w of length n contains at most $n+1$ distinct palindromic factors.

Every factor of a rich word is also rich.

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).

Properties Rich words I

A finite word w of length n contains at most $n + 1$ distinct palindromic factors.

Every factor of a rich word is also rich.

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).

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

Properties Rich words II

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.

Properties Rich words II

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 characteristic property of rich words is that all complete returns to any palindromic factor u in w are palindromes.

Properties Rich words II

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 characteristic property of rich words is that all complete returns to any palindromic factor u in w are palindromes.

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

Properties Rich words III

A factor *r* of a rich word *w* is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.

[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.]

Properties Rich words III

A factor r of a rich word w is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.

[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.]

Example

There is no rich word containing both 010110110 and 01011110110 as factors.

Enumeration of Rich Words - Lower and Upper bounds I

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.

Enumeration of Rich Words - Lower and Upper bounds I

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.

J. VESTI, *Extensions of rich words*, Theor. Comput. Sci., 548 (2014), pp. 14–24.

Enumeration of Rich Words - Lower and Upper bounds I

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.

J. VESTI, *Extensions of rich words*, Theor. Comput. Sci., 548 (2014), pp. 14–24.

Let $R(n)$ denote the number of rich words of length over a given finite alphabet.

Enumeration of Rich Words - Lower and Upper bounds I

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.

J. VESTI, *Extensions of rich words*, Theor. Comput. Sci., 548 (2014), pp. 14–24.

Let $R(n)$ denote the number of rich words of length over a given finite alphabet.

C. Guo, J. Shallit and A.M. Shur constructed 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 $R(n) \geq \frac{C^{\sqrt{n}}}{p(n)}$, where $p(n)$ is a polynomial and the constant $C \approx 37$.

Enumeration of Rich Words - Lower and Upper bounds II

C. Guo, J. Shallit and A.M. Shur used the calculation performed by M. Rubinchik to provide an exponencial upper bound for binary rich words $R(n) \leq c1.605^n$, where is c some constant.

Enumeration of Rich Words - Lower and Upper bounds II

C. Guo, J. Shallit and A.M. Shur used the calculation performed by M. Rubinchik to provide an exponencial upper bound for binary rich words $R(n) \leq c1.605^n$, where is c some constant.

Guo, Shallit and Shur conjectured that the number of rich words grows slightly slower than $n^{\sqrt{n}}$.

Enumeration of Rich Words - Lower and Upper bounds II

C. Guo, J. Shallit and A.M. Shur used the calculation performed by M. Rubinchik to provide an exponencial upper bound for binary rich words $R(n) \leq c1.605^n$, where is c some constant.

Guo, Shallit and Shur conjectured that the number of rich words grows slightly slower than $n^{\sqrt{n}}$.

C. GUO, J. SHALLIT, AND A. M. SHUR, *Palindromic rich words and run-length encodings*, Inform. Process. Lett., 116 (2016), pp. 735–738.

Palindromic length

A *palindromic length* $PL(v)$ of a finite word v is the minimal number of palindromes whose concatenation is equal to v.

Palindromic length

A *palindromic length* $PL(v)$ of a finite word v is the minimal number of palindromes whose concatenation is equal to v.

It was conjectured in 2013 that for every aperiodic infinite word x, the palindromic length of its factors is not bounded.

Palindromic length

A *palindromic length* $PL(v)$ of a finite word v is the minimal number of palindromes whose concatenation is equal to v.

It was conjectured in 2013 that for every aperiodic infinite word x, the palindromic length of its factors is not bounded.

A. FRID, S. PUZYNINA, AND L. ZAMBONI, *On palindromic factorization of words*, Adv. Appl. Math., 50 (2013), pp. 737–748.

On the Number of Rich Words

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.

On the Number of Rich Words

Lemma

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

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

On the Number of Rich Words

Lemma

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

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

On the Number of Rich Words

Lemma

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

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

In general, for non-rich words, UPS-factorization does not need to exist.

On the Number of Rich Words

Theorem

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*

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

On the Number of Rich Words

Theorem

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

$$p \leq c\frac{n}{\ln n}. \tag{2}$$

The theorem says that a rich word w is concatenated from a "small" number of palindromes. "Small" means that $\lim_{n\to\infty} \frac{c\frac{n}{\ln n}}{n} = 0$.

On the Number of Rich Words

Realize that $\sum_{i=1}^{t} iq^{\lceil \frac{i}{2} \rceil}$ is the length of the word, which is constructed as a concatenation of all palindromes of length $\leq t$.

On the Number of Rich Words

Realize that $\sum_{i=1}^{t} iq^{\lceil \frac{i}{2} \rceil}$ is the length of the word, which is constructed as a concatenation of all palindromes of length $\leq t$.

Lemma

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

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

On the Number of Rich Words

Let us define

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

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

On the Number of Rich Words

Let us define

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

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

Theorem

If $n \geq 2$, then

$$R(n) \leq \sum_{p=1}^{\kappa_n} \sum_{\substack{n_1+n_2+\cdots+n_p=n \\ n_1,n_2,\ldots,n_p \geq 1}} R\left(\left\lceil \frac{n_1}{2} \right\rceil\right) R\left(\left\lceil \frac{n_2}{2} \right\rceil\right) \ldots R\left(\left\lceil \frac{n_p}{2} \right\rceil\right) . \quad (5)$$

On the Number of Rich Words

Theorem

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

An Upper Bound for Palindromic and Factor Complexity of Rich Words

RUKAVICKA, JOSEF, *Upper bound for palindromic and factor complexity of rich words*, RAIRO-Theor. Inf. Appl., 55 (2021) Article No. 1.

An Upper Bound for Palindromic and Factor Complexity of Rich Words

Let $F(w, n)$ be the set of factors of length n of the word w and let $F_p(w, n) \subseteq F(w, n)$ be the set of palindromic factors.

An Upper Bound for Palindromic and Factor Complexity of Rich Words

Let $F(w,n)$ be the set of factors of length n of the word w and let $F_p(w,n) \subseteq F(w,n)$ be the set of palindromic factors.

Given a palindrome u and $a, b \in A$, where $a \neq b$. We call the word aub a u-switch.

An Upper Bound for Palindromic and Factor Complexity of Rich Words

Let $F(w, n)$ be the set of factors of length n of the word w and let $F_p(w, n) \subseteq F(w, n)$ be the set of palindromic factors.

Given a palindrome u and $a, b \in A$, where $a \neq b$. We call the word aub a u-switch.

Theorem

If $xux, yuy \in F_p(w, |u| + 2)$, where $x, y \in A$ and $x \neq y$, then w contains a u-switch, formally there is $aub \in F(w, |u| + 2)$.

An Upper Bound for Palindromic and Factor Complexity of Rich Words

Let R denote the set of rich words (both finite and infinite). Let $F(w) = \bigcup_{j \geq 0} F(w, j)$ and $F_p(w) = \bigcup_{j \geq 0} F_p(w, j)$. Let $lpps(w)$ be the longest proper palindromic suffix of the word w.

An Upper Bound for Palindromic and Factor Complexity of Rich Words

Let R denote the set of rich words (both finite and infinite). Let $F(w) = \bigcup_{j \geq 0} F(w, j)$ and $F_p(w) = \bigcup_{j \geq 0} F_p(w, j)$. Let $lpps(w)$ be the longest proper palindromic suffix of the word w.

Theorem

Let $w \in R$, $u, v \in F_p(w)$, $lpps(u) = lpps(v)$, $a, b \in A$ and $a \neq b$. Then $aub, avb \in F(w)$ implies that $u = v$.

An Upper Bound for Palindromic and Factor Complexity of Rich Words

We prove a quasi-polynomial upper bound for the palindromic and factor complexity of rich words. Let $\delta = \frac{3}{2(\ln 3 - \ln 2)}$.

Theorem

If $w \in RW \cup RW^{\infty}$ and $n \in \mathbb{N}_1$, then

$$|F(w,n)| \leq (4q^2 n)^{\delta \ln 2n + 2}.$$

Palindromic factorization of rich words

RUKAVICKA, JOSEF, *Palindromic factorization of rich words*, Discrete Applied Mathematics, Volume 316, 2022, Pages 95-102.

Palindromic factorization of rich words

RUKAVICKA, JOSEF, *Palindromic factorization of rich words*, Discrete Applied Mathematics, Volume 316, 2022, Pages 95-102.

Let $LUF(w) = p$ be *the length of UPS-factorization* of w.

Palindromic factorization of rich words

RUKAVICKA, JOSEF, *Palindromic factorization of rich words*, Discrete Applied Mathematics, Volume 316, 2022, Pages 95-102.

Let $LUF(w) = p$ be *the length of UPS-factorization* of w.

Theorem

For a given finite alphabet $\mathcal{A}$, there are real positive constants μ, κ such that, if w is a finite nonempty rich word over the alphabet $\mathcal{A}$ and $n = |w|$, then

$$LUF(w) \leq \mu \frac{n}{e^{\kappa\sqrt{\ln n}}}.$$

Palindromic factorization of rich words

We conjecture that:
For a given finite alphabet $\mathcal{A}$, there is a positive real constant λ such that, if w is a finite nonempty rich word over the alphabet $\mathcal{A}$ and $n = |w|$, then $LUF(w) \leq \lambda\sqrt{n}$.

Palindromic factorization of rich words

We conjecture that:
For a given finite alphabet $\mathcal{A}$, there is a positive real constant λ such that, if w is a finite nonempty rich word over the alphabet $\mathcal{A}$ and $n = |w|$, then $LUF(w) \leq \lambda\sqrt{n}$.

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

$$c_1 > \max\{e^{(\delta \ln 2+2)(\ln 4+2\ln q)}, e^{-(\beta+g\beta+c_2\beta^2)}\}.$$

The constant g, where $0 < g < 1$, can be chosen arbitrarily.

Palindromic factorization of rich words

From Theorem 12

Corollary

If $w \in RW \cup RW^{\infty}$ and $n \in \mathbb{N}_1$, then we have that

$$|F(w,n)| \leq c_1 n^{c_2 \ln n}.$$

Palindromic factorization of rich words

The sum $\sum_{i=1}^{k} i\lfloor c_1 i^{c_2 \ln i} \rfloor$ has the following interpretation: It is the length of the word w which is concatenation of $\lfloor c_1 i^{c_2 \ln i} \rfloor$ words of length i for $i \in \{1, 2, \ldots, k\}$.

Palindromic factorization of rich words

The sum $\sum_{i=1}^{k} i\lfloor c_1 i^{c_2 \ln i}\rfloor$ has the following interpretation: It is the length of the word w which is concatenation of $\lfloor c_1 i^{c_2 \ln i}\rfloor$ words of length i for $i \in \{1, 2, \ldots, k\}$.

Lemma

There is $k_0 \in \mathbb{N}_1$ such that for $k \geq k_0$, we have that

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

Palindromic factorization of rich words

Lemma

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

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

Palindromic factorization of rich words

Lemma

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

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

Corollary

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

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

Palindromic factorization of rich words

Let $\gamma = \beta + \ln c_1 + c_2\beta^2$.

Lemma

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

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

Palindromic factorization of rich words

Lemma

If $\bar{t}, t \in \mathbb{N}_1$, $\bar{t} > t$, $\overline{\omega}, \omega : \mathbb{N}_1 \to \mathbb{N}_1$, $\overline{\omega}(i) \leq \omega(i)$ *for every* $i \in \mathbb{N}_1$, *and*

$$\sum_{i=1}^{\bar{t}} \overline{\omega}(i) > \sum_{i=1}^{t} \omega(i),$$

then

$$\sum_{i=1}^{\bar{t}} i\overline{\omega}(i) > \sum_{i=1}^{t} i\omega(i).$$

Palindromic factorization of rich words

Let $\overline{\Omega}, \Omega \subset \mathcal{A}^+$ be two finite sets. Let $\overline{\omega}(i) = |\overline{\Omega} \cap \mathcal{A}^i|$ and $\omega(i) = |\Omega \cap \mathcal{A}^i|$. Let $\overline{w}, w \in \mathcal{A}^+$ be words that are a concatenation of all words from $\overline{\Omega}, \Omega$ respectively. The previous Lemma implies that: if $|\overline{\Omega}| > |\Omega|$ and $\overline{\omega}(i) \leq \omega(i)$, then $|\overline{w}| > |w|$.

Palindromic factorization of rich words

Let $\overline{\Omega}, \Omega \subset \mathcal{A}^+$ be two finite sets. Let $\overline{\omega}(i) = |\overline{\Omega} \cap \mathcal{A}^i|$ and $\omega(i) = |\Omega \cap \mathcal{A}^i|$. Let $\overline{w}, w \in \mathcal{A}^+$ be words that are a concatenation of all words from $\overline{\Omega}, \Omega$ respectively. The previous Lemma implies that: if $|\overline{\Omega}| > |\Omega|$ and $\overline{\omega}(i) \leq \omega(i)$, then $|\overline{w}| > |w|$.

Theorem

If $w \in RW \cap \mathcal{A}^+$, $n = |w|$, $t \in \mathbb{N}_1$, and

$$\sum_{i=1}^{t} i \lfloor c_1 i^{c_2 \ln i} \rfloor \geq n,$$

then

$$LUF(w) \leq \sum_{i=1}^{t} c_1 i^{c_2 \ln i}.$$

Rich words - Open question

- Improving of the upper bound for the number of rich words. Using our result for palindromic complexity, we expect to prove that $R(n) \leq q^{c(q)\frac{n}{2^{\sqrt{\ln n}}}}$, where $n > 1$, $q > 1$ and $c(q)$ is a constant depending on q.

Rich words - Open question

- Improving of the upper bound for the number of rich words. Using our result for palindromic complexity, we expect to prove that $R(n) \leq q^{c(q)\frac{n}{2^{\sqrt{\ln n}}}}$, where $n > 1$, $q > 1$ and $c(q)$ is a constant depending on q.

- Is there a polynomial upper bound for the palindromic complexity of rich words: $F(w, n) \leq n^{c(q)}$, where $n > 1$.

Thank you