



FACTOR FREQUENCIES IN LANGUAGES INVARIANT UNDER SYMMETRIES PRESERVING FACTOR FREQUENCIES

Ľubomíra Balková

Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic
lubomira.balkova@fjfi.cvut.cz

Received: 12/3/11, Accepted: 5/30/12, Published: 6/25/12

Abstract

The number of frequencies of factors of length $n + 1$ in a recurrent aperiodic infinite word does not exceed $3\Delta\mathcal{C}(n)$, where $\Delta\mathcal{C}(n)$ is the first difference of factor complexity, as shown by Boshernitzan. Pelantová together with the author derived a better upper bound for infinite words whose language is closed under reversal. In this paper, we further diminish the upper bound for uniformly recurrent infinite words whose language is invariant under all elements of a finite group of symmetries and we prove the optimality of the obtained upper bound.

1. Introduction

When studying factor frequencies, the Rauzy graph is a powerful tool. Using this tool, the following results were obtained. Dekking in [8] described factor frequencies of two famous infinite words – the Fibonacci word and the Thue-Morse word. Using Rauzy graphs, it is readily seen that frequencies of factors of a given length of any Arnoux-Rauzy word over an m -letter alphabet take at most $m + 1$ distinct values. Explicit values of factor frequencies were derived by Berthé in [4] for Sturmian words and by Wozny and Zamboni in [17] for Arnoux-Rauzy words in general.

Queffélec in [15] explored factor frequencies of fixed points of morphisms from another point of view – as a shift invariant probability measure. She provided a rather complicated algorithm for the computation of values of such a measure. For some special classes of fixed points of morphisms (circular marked uniform morphisms), Frid [10] described explicitly their factor frequencies.

A simple idea concerning Rauzy graphs lead Boshernitzan [5] to an upper bound on the number of different factor frequencies in an arbitrary recurrent aperiodic infinite word. He showed that the number of frequencies of factors of length $n + 1$ does not exceed $3\Delta\mathcal{C}(n)$, where $\Delta\mathcal{C}(n)$ is the first difference of factor complexity. In [6], it was shown that $\Delta\mathcal{C}(n)$ is bounded for infinite words with sublinear complexity

(for instance, fixed points of primitive substitutions is a subclass of infinite words with sublinear complexity), therefore the number of different frequencies of factors of the same length is bounded.

In our previous paper [2], making use of reflection symmetry of Rauzy graphs, we diminished Boshernitzan's upper bound for infinite words whose language is closed under reversal.

This time, we generalize our result to infinite words whose language is invariant under all elements of a group of symmetries and whose Rauzy graphs are therefore invariant under all elements of a group of automorphisms. In Section 2, we introduce basic notions, describe the main tool of our proofs – reduced Rauzy graphs – and summarize in detail the known upper bounds on the number of factor frequencies. In Section 3, we define symmetries preserving factor frequencies and show how all such symmetries look like. In Section 4, we prove Theorem 20, which provides an optimal upper bound on the number of factor frequencies of infinite words whose language is invariant under all elements of a finite group of symmetries. Section 5 is devoted to the demonstration that the upper bound from the main theorem is indeed optimal.

Finally, let us mention that the idea to exploit symmetries of the Rauzy graph was already used in [3] in order to estimate the number of palindromes of a given length, and, recently, it has been used profoundly in [12, 13, 16] for the generalization of the so-called rich and almost rich words (see [11]) for languages invariant under more symmetries than just reversal.

2. Preliminaries

An *alphabet* \mathcal{A} is a finite set of symbols, called *letters*. A concatenation of letters is a *word*. The *length* of a word w is the number of letters in w and is denoted $|w|$. The set \mathcal{A}^* of all finite words (including the empty word ε) provided with the operation of concatenation is a free monoid. The set of all finite words but the empty word ε is denoted \mathcal{A}^+ . We will also deal with right-sided infinite words $\mathbf{u} = u_0u_1u_2\dots$, where $u_i \in \mathcal{A}$. A finite word w is called a *factor* of the word u (finite or infinite) if there exist a finite word p and a word s (finite or infinite) such that $u = pws$. The factor p is a *prefix* of u and s is a *suffix* of u . An infinite word \mathbf{u} is said to be *recurrent* if each of its factors occurs infinitely many times in \mathbf{u} . An *occurrence* of a finite word w in a finite word $v = v_1v_2\dots v_m$ (in an infinite word \mathbf{u}) is an index i such that w is a prefix of the word $v_iv_{i+1}\dots v_m$ (of the word $u_iu_{i+1}\dots$). An infinite word \mathbf{u} is called *uniformly recurrent* if for any factor w the set $\{j - i \mid i \text{ and } j \text{ are successive occurrences of } w \text{ in } \mathbf{u}\}$ is bounded.

2.1. Complexity and Special Factors

The *language* $\mathcal{L}(\mathbf{u})$ of an infinite word \mathbf{u} is the set of all factors of \mathbf{u} . Let $\mathcal{L}_n(\mathbf{u})$ denote the set of factors of length n of \mathbf{u} . The *factor complexity* (or *complexity*) of \mathbf{u} is the mapping $\mathcal{C} : \mathbb{N} \rightarrow \mathbb{N}$ which associates with every n the number of different factors of length n of \mathbf{u} , i.e., $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$.

An important role for the computation of factor complexity is played by special factors. We say that a letter a is a *right extension* of a factor $w \in \mathcal{L}(\mathbf{u})$ if wa is also a factor of \mathbf{u} . Let $\text{Rext}(w)$ denote the set of all right extensions of w in \mathbf{u} , i.e., $\text{Rext}(w) = \{a \in \mathcal{A} \mid wa \in \mathcal{L}(\mathbf{u})\}$. If $\#\text{Rext}(w) \geq 2$, then the factor w is called *right special* (RS for short). Analogously, we define *left extensions*, $\text{Lext}(w)$, *left special factors* (LS for short). Moreover, we say that a factor w is *bispecial* (BS for short) if w is LS and RS.

With these notions in hand, we may introduce a formula for the *first difference of complexity* $\Delta\mathcal{C}(n) = \mathcal{C}(n + 1) - \mathcal{C}(n)$ (taken from [7]).

$$\Delta\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} (\#\text{Rext}(w) - 1) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} (\#\text{Lext}(w) - 1), \quad n \in \mathbb{N}. \quad (1)$$

2.2. Morphisms and Antimorphisms

A mapping φ on \mathcal{A}^* is called

- a *morphism* if $\varphi(vw) = \varphi(v)\varphi(w)$ for any $v, w \in \mathcal{A}^*$,
- an *antimorphism* if $\varphi(vw) = \varphi(w)\varphi(v)$ for any $v, w \in \mathcal{A}^*$.

Let $AM(\mathcal{A}^*)$ denote the set of all morphisms and antimorphisms on \mathcal{A}^* . Together with composition, it forms a monoid (the unit element is the identity mapping Id). The *mirror* (also called *reversal*) *mapping* R defined by $R(w_1w_2 \dots w_{m-1}w_m) = w_mw_{m-1} \dots w_2w_1$ is an involutive antimorphism, i.e., $R^2 = \text{Id}$. It is obvious that any antimorphism is a composition of R and a morphism.

A language $\mathcal{L}(\mathbf{u})$ is *closed (invariant) under reversal* if for every factor $w \in \mathcal{L}(\mathbf{u})$, also its *mirror image* $R(w)$ belongs to $\mathcal{L}(\mathbf{u})$. A factor w which coincides with its mirror image $R(w)$ is called a *palindrome*. More generally, a language $\mathcal{L}(\mathbf{u})$ is *closed (invariant) under an antimorphism or morphism* $\Psi \in AM(\mathcal{A}^*)$ if for every factor $w \in \mathcal{L}(\mathbf{u})$, $\Psi(w)$ also belongs to $\mathcal{L}(\mathbf{u})$. If θ is an antimorphism on \mathcal{A}^* , then any word w such that $w = \theta(w)$ is called a θ -*palindrome*. It is not difficult to see that an infinite word whose language is closed under an antimorphism θ of finite order, i.e., such that there exists $k \in \mathbb{N}, k \geq 1$, satisfying $\theta^k = \text{Id}$, is recurrent.

We define the θ -*palindromic complexity* of the infinite word \mathbf{u} as the mapping $\mathcal{P}_\theta : \mathbb{N} \rightarrow \mathbb{N}$ satisfying $\mathcal{P}_\theta(n) = \#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w = \theta(w)\}$. If $\theta = R$, we write $\mathcal{P}(n)$ instead of $\mathcal{P}_R(n)$. Clearly, $\mathcal{P}(n) \leq \mathcal{C}(n)$ for all $n \in \mathbb{N}$. A non-trivial inequality between $\mathcal{P}(n)$ and $\mathcal{C}(n)$ can be found in [1]. Here, we use a result from [3].

Theorem 1. *If the language of an infinite word is closed under reversal, then for all $n \in \mathbb{N}$, we have*

$$\mathcal{P}(n) + \mathcal{P}(n + 1) \leq \Delta\mathcal{C}(n) + 2. \tag{2}$$

This result has been recently generalized in [13].

Theorem 2. *Let $G \subset AM(\mathcal{A}^*)$ be a finite group containing an antimorphism and let \mathbf{u} be an infinite word whose language is invariant under all elements of G . If there exists an integer $N \in \mathbb{N}$ such that any factor of \mathbf{u} of length N contains all letters of \mathcal{A} , then*

$$\sum_{\theta \in G^{(2)}} (\mathcal{P}_\theta(n) + \mathcal{P}_\theta(n + 1)) \leq \Delta\mathcal{C}(n) + \#G \quad \text{for all } n \geq N,$$

where $G^{(2)}$ is the set of involutive antimorphisms in G .

In fact, even a weaker assumption on N in Theorem 2 may be imposed.

Definition 3. Let $G \subset AM(\mathcal{A}^*)$ be a finite group containing an antimorphism and let \mathbf{u} be an infinite word whose language is invariant under all elements of G . We say that a number $N \in \mathbb{N}$ is G -distinguishing on \mathbf{u} if for any $w \in \mathcal{L}_N(\mathbf{u})$ the following holds:

$$\theta_1 \neq \theta_2 \Rightarrow \theta_1(w) \neq \theta_2(w) \quad \text{for any two antimorphisms } \theta_1, \theta_2 \in G.$$

As explained in [14], instead of the assumption on N from Theorem 2, it suffices to assume that N is a G -distinguishing number on \mathbf{u} .

Remark 4. If \mathbf{u} is an infinite word whose language is closed under reversal, i.e., invariant under a morphism and an antimorphism of $G = \{\text{Id}, R\}$, then any number $N \geq 1$ is G -distinguishing, and hence the inequality from Theorem 2 holds for $n \geq 1$. Moreover, the inequality holds also for $n = 0$. Consequently, Theorem 1 is indeed a particular case of Theorem 2.

2.3. Factor Frequency

If w is a factor of an infinite word \mathbf{u} and if

$$\lim_{|v| \rightarrow \infty, v \in \mathcal{L}(\mathbf{u})} \frac{\#\{\text{occurrences of } w \text{ in } v\}}{|v|}$$

exists, then it is denoted by $\rho(w)$ and called the *frequency* of w .

Let us recall a result of Frid [10], which is useful for the calculation of factor frequencies in fixed points of primitive morphisms. In order to introduce the result, we need some further notions. Let φ be a morphism on $\mathcal{A}^* = \{a_1, a_2, \dots, a_m\}^*$. We associate with φ the *incidence matrix* M_φ given by $[M_\varphi]_{ij} = |\varphi(a_j)|_{a_i}$, where

$|\varphi(a_j)|_{a_i}$ denotes the number of occurrences of a_i in $\varphi(a_j)$. The morphism φ is called *primitive* if there exists $k \in \mathbb{N}$ satisfying that the power M_φ^k has all entries strictly positive. As shown in [15], for fixed points of primitive morphisms,

- factor frequencies exist,
- it follows from the Perron-Frobenius theorem that the incidence matrix has one dominant eigenvalue λ , which is larger than the modulus of any other eigenvalue,
- the components of the unique eigenvector $(x_1, x_2, \dots, x_m)^T$ corresponding to λ normalized so that $\sum_{i=1}^m x_i = 1$ coincide with the letter frequencies, i.e., $x_i = \rho(a_i)$ for all $i \in \{1, 2, \dots, m\}$.

Let φ be a morphism on \mathcal{A}^* . Let $\psi_{ij} : \mathcal{A}^+ \rightarrow \mathcal{A}^+$, where $i, j \in \mathbb{N}$, denote the mapping that associates with $v \in \mathcal{A}^+$ the word $\psi_{ij}(v)$ obtained from $\varphi(v)$ by erasing i letters from the left and j letters from the right, where $i + j < |\varphi(v)|$. We say that a word $v \in \mathcal{A}^+$ admits an *interpretation* $s = (b_0 b_1 \dots b_m, i, j)$ if $v = \psi_{ij}(b_0 b_1 \dots b_m)$, where $b_i \in \mathcal{A}$ and $i < |\varphi(b_0)|$ and $j < |\varphi(b_m)|$. The word $a(s) = b_0 b_1 \dots b_m$ is an *ancestor* of s . The set of all interpretations of v is denoted $I(v)$. Now we can recall the promised result of Frid [10].

Proposition 5. *Let φ be a primitive morphism having a fixed point \mathbf{u} and let λ be the dominant eigenvalue of the incidence matrix M_φ . Then for any factor $v \in \mathcal{L}(\mathbf{u})$, we have*

$$\rho(v) = \frac{1}{\lambda} \sum_{s \in I(v)} \rho(a(s)).$$

2.4. Reduced Rauzy Graphs

Assume throughout this section that factor frequencies of infinite words in question exist. The *Rauzy graph* of order n of an infinite word \mathbf{u} is a directed graph Γ_n whose set of vertices is $\mathcal{L}_n(\mathbf{u})$ and set of edges is $\mathcal{L}_{n+1}(\mathbf{u})$. An edge $e = w_0 w_1 \dots w_n$ starts in the vertex $w = w_0 w_1 \dots w_{n-1}$, ends in the vertex $v = w_1 \dots w_{n-1} w_n$, and is labeled by its factor frequency $\rho(e)$.

Edge frequencies in a Rauzy graph Γ_n behave similarly as the current in a circuit. We may formulate an analogy of the Kirchhoff current law: the sum of frequencies of edges ending in a vertex equals the sum of frequencies of edges starting in this vertex.

Observation 6. *(Kirchhoff law for frequencies) Let w be a factor of an infinite word \mathbf{u} whose factor frequencies exist. Then*

$$\rho(w) = \sum_{a \in \text{Lext}(w)} \rho(aw) = \sum_{a \in \text{Rext}(w)} \rho(wa).$$

The Kirchhoff law for frequencies has some useful consequences.

Corollary 7. *Let w be a factor of an infinite word \mathbf{u} whose frequency exists.*

- *If w has a unique right extension a , then $\rho(w) = \rho(wa)$.*
- *If w has a unique left extension a , then $\rho(w) = \rho(aw)$.*

Corollary 8. *Let w be a factor of an aperiodic recurrent infinite word \mathbf{u} whose frequency exists. Let v be the shortest BS factor containing w , then $\rho(w) = \rho(v)$.*

The assumption of recurrence and aperiodicity in Corollary 8 is needed in order to ensure that every factor can be extended to a BS factor.

Corollary 7 implies that if a Rauzy graph contains a vertex w with only one incoming edge aw and one outgoing edge wb , then $\rho := \rho(aw) = \rho(w) = \rho(wb) = \rho(awb)$. Therefore, we can replace this triplet (edge-vertex-edge) with only one edge awb keeping the frequency ρ . If we reduce the Rauzy graph step by step applying the above described procedure, we obtain the so-called *reduced Rauzy graph* $\tilde{\Gamma}_n$, which simplifies the investigation of edge frequencies. In order to make this construction precise this construction, we introduce the notion of a simple path.

Definition 9. Let Γ_n be the Rauzy graph of order n of an infinite word \mathbf{u} . A factor e of length larger than n such that its prefix and its suffix of length n are special factors and e does not contain any other special factors is called a simple path. We define the label of a simple path e as $\rho(e)$.

Definition 10. The reduced Rauzy graph $\tilde{\Gamma}_n$ of \mathbf{u} of order n is a directed graph whose set of vertices is formed by LS and RS factors of $\mathcal{L}_n(\mathbf{u})$ and whose set of edges is given in the following way. Vertices w and v are connected with an edge e if there exists in Γ_n a simple path starting in w and ending in v . We assign to such an edge e the label of the corresponding simple path.

For a recurrent word \mathbf{u} , at least one edge starts and at least one edge ends in every vertex of Γ_n . If \mathbf{u} is moreover aperiodic, then all its Rauzy graphs contain at least one LS and one RS factor. It is thus not difficult to see that for recurrent aperiodic words, the set of edge labels in Γ_n is equal to the set of edge labels in the reduced Rauzy graph $\tilde{\Gamma}_n$. The number of edge labels in the Rauzy graph $\tilde{\Gamma}_n$ is clearly less than or equal to the number of edges in $\tilde{\Gamma}_n$. Let us calculate the number of edges in $\tilde{\Gamma}_n$ in order to get an upper bound on the number of frequencies of factors in $\mathcal{L}_{n+1}(\mathbf{u})$.

For every RS factor $w \in \mathcal{L}_n(\mathbf{u})$, we have that $\#\text{Rext}(w)$ edges begin in w and for every LS factor $v \in \mathcal{L}_n(\mathbf{u})$ which is not RS, only one edge begins in v , thus we get the following formula

$$\#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\} = \sum_{w \text{ RS in } \mathcal{L}_n(\mathbf{u})} \#\text{Rext}(w) + \sum_{v \text{ LS not RS in } \mathcal{L}_n(\mathbf{u})} 1. \quad (3)$$

We rewrite the first term using (1) and the second term using the definition of BS factors in the following way

$$\#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\} = \Delta C(n) + \sum_{v \text{ RS in } \mathcal{L}_n(\mathbf{u})} 1 + \sum_{v \text{ LS in } \mathcal{L}_n(\mathbf{u})} 1 - \sum_{v \text{ BS in } \mathcal{L}_n(\mathbf{u})} 1. \quad (4)$$

Since $\#\text{Rext}(w) - 1 \geq 1$ for any RS factor w and, similarly, for LS factors, we have

$$\#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w \text{ RS}\} \leq \Delta C(n) \quad \text{and} \quad \#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w \text{ LS}\} \leq \Delta C(n). \quad (5)$$

By combining (4) and (5), we obtain

$$\#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\} \leq 3\Delta C(n) - X, \quad (6)$$

where X is the number of BS factors of length n . This provides us with the result already proved by Boshernitzan in [5].

Theorem 11. *Let \mathbf{u} be an aperiodic recurrent infinite word such that the frequency $\rho(w)$ exists for every factor $w \in \mathcal{L}(\mathbf{u})$. Then for every $n \in \mathbb{N}$, we have*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \leq 3\Delta C(n).$$

In the paper [2], we considered infinite words with language closed under reversal and we lowered the upper bound from Theorem 11 for them.

Theorem 12. *Let \mathbf{u} be an infinite word whose language $\mathcal{L}(\mathbf{u})$ is closed under reversal and such that the frequency $\rho(w)$ exists for every factor $w \in \mathcal{L}(\mathbf{u})$. Then for every $n \in \mathbb{N}$, we have*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \leq 2\Delta C(n) + 1 - \frac{1}{2}X - \frac{1}{2}Y, \quad (7)$$

where X is the number of BS factors of length n and Y is the number of palindromic BS factors of length n .

Corollary 13. *Let \mathbf{u} be an infinite word whose language $\mathcal{L}(\mathbf{u})$ is closed under reversal and such that the frequency $\rho(w)$ exists for every factor $w \in \mathcal{L}(\mathbf{u})$. Then the number of distinct factor frequencies obeys for all $n \in \mathbb{N}$,*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \leq 2\Delta C(n) + 1, \quad (8)$$

where the equality is reached if and only if \mathbf{u} is purely periodic.

As shown by Ferenczi and Zamboni [9], m -iet words attain the upper bound from (7) for all $n \in \mathbb{N}$. Since Sturmian words are 2-iet words, they reach the upper bound in (7) for all $n \in \mathbb{N}$, too. Consequently, the upper bound from (7) is optimal and cannot be improved while preserving the assumptions. However, as we will show in the sequel, if the language of an infinite word \mathbf{u} is invariant under more symmetries, the upper bound from (7) may be lowered considerably.

3. Symmetries Preserving Factor Frequency

We will be interested in symmetries preserving in a certain way factor occurrences in \mathbf{u} , and consequently, frequencies of factors of \mathbf{u} . Let us call a *symmetry* on \mathcal{A}^* any mapping Ψ satisfying the following two properties:

1. Ψ is a bijection: $\mathcal{A}^* \rightarrow \mathcal{A}^*$,
2. for all $w, v \in \mathcal{A}^*$

$$\#\{\text{occurrences of } w \text{ in } v\} = \#\{\text{occurrences of } \Psi(w) \text{ in } \Psi(v)\}.$$

Theorem 14. *Let $\Psi : \mathcal{A}^* \rightarrow \mathcal{A}^*$. Then Ψ is a symmetry if and only if Ψ is a morphism or an antimorphism such that Ψ is a letter permutation when restricted to \mathcal{A} .*

The proof of Theorem 14 is obtained when putting together the following two lemmas.

Lemma 15. *Let Ψ be a symmetry on \mathcal{A}^* and let $w \in \mathcal{A}^*$. Then $|\Psi(w)| = |w|$.*

Proof. Since $\#\{\text{occurrences of } \Psi(w) \text{ in } \Psi(\varepsilon)\} = \#\{\text{occurrences of } w \text{ in } \varepsilon\} = 0$ for every $w \in \mathcal{A}^*$, it follows that $\Psi(\varepsilon) = \varepsilon$.

Since Ψ is a bijection, for every letter $a \in \mathcal{A}$, there exists a unique $w \in \mathcal{A}^*$ such that $\Psi(w) = a$, where $w \neq \varepsilon$. Hence, for any letter b , we have $\#\{\text{occurrences of } b \text{ in } w\} = \#\{\text{occurrences of } \Psi(b) \text{ in } a\}$. Consequently, since Ψ is a bijection, $|w| = 1$. If we denote $\mathcal{A} = \{a_1, \dots, a_m\}$, then it follows that there exists a permutation $\pi \in S_m$ such that $\Psi(a_k) = a_{\pi(k)}$ for all $k \in \{1, \dots, m\}$.

Let us now take an arbitrary $w \in \mathcal{A}^*$, then using the fact that Ψ restricted to \mathcal{A} is a letter permutation and applying Property (2), we have

$$|w| = \sum_{a \in \mathcal{A}} \#\{\text{occurrences of } a \text{ in } w\} = \sum_{a \in \mathcal{A}} \#\{\text{occurrences of } \Psi(a) \text{ in } \Psi(w)\} = |\Psi(w)|.$$

□

Corollary 16. *For every $w_1 w_2 \dots w_n \in \mathcal{A}^*$, $w_i \in \mathcal{A}$, the following equation is valid*

$$\Psi(w_1 w_2 \dots w_n) = \Psi(w_{\sigma(1)}) \Psi(w_{\sigma(2)}) \dots \Psi(w_{\sigma(n)}) \tag{9}$$

for some permutation $\sigma \in S_n$.

The next lemma claims that the permutation σ is necessarily either the identical permutation $(1 \ 2 \ \dots \ n)$ or the symmetric permutation $(n \ \dots \ 2 \ 1)$.

Lemma 17. *Let Ψ be a symmetry on \mathcal{A}^* . Then Ψ is either a morphism or an antimorphism.*

Proof. We have to show that $\Psi(w) = \Psi(w_1)\Psi(w_2) \dots \Psi(w_n)$ for every $w = w_1w_2 \dots w_n \in \mathcal{A}^*$, $w_i \in \mathcal{A}$, or $\Psi(w) = \Psi(w_n) \dots \Psi(w_2)\Psi(w_1)$ for every $w = w_1w_2 \dots w_n \in \mathcal{A}^*$, $w_i \in \mathcal{A}$.

Let us proceed by induction on the length n of w . The case $n = 1$ is clear. Suppose that $\Psi(w) = \Psi(w_1)\Psi(w_2) \dots \Psi(w_{n-1})$ for every $w = w_1w_2 \dots w_{n-1} \in \mathcal{A}^*$ of length $n - 1$, $n \geq 2$. Take an arbitrary word $w = w_1w_2 \dots w_n \in \mathcal{A}^*$. Then, as Ψ is a symmetry, $\Psi(w_2 \dots w_n)$ is a factor of $\Psi(w_1w_2 \dots w_n)$, in more precise terms, $\Psi(w_2 \dots w_n)$ is either a prefix or a suffix of $\Psi(w_1 \dots w_n)$. Moreover, if w_1 occurs in $w_2 \dots w_n$ ℓ times, w_1 occurs in $w_1w_2 \dots w_n$ $(\ell + 1)$ times. Since Ψ is a symmetry, it follows that $\Psi(w_1)$ occurs ℓ times in $\Psi(w_2 \dots w_n)$ and $(\ell + 1)$ times in $\Psi(w_1w_2 \dots w_n)$. These two observations result in

$$\Psi(w_1w_2 \dots w_n) = \Psi(w_1)\Psi(w_2 \dots w_n) = \Psi(w_1)\Psi(w_2) \dots \Psi(w_n)$$

or

$$\Psi(w_1w_2 \dots w_n) = \Psi(w_2 \dots w_n)\Psi(w_1) = \Psi(w_2) \dots \Psi(w_n)\Psi(w_1).$$

The first case is in correspondence with the fact that Ψ is a morphism. Let us treat the second case. Similar reasoning as before leads to

$$\Psi(w_1w_2 \dots w_n) = \Psi(w_1 \dots w_{n-1})\Psi(w_n) = \Psi(w_1)\Psi(w_2) \dots \Psi(w_n)$$

or

$$\Psi(w_1w_2 \dots w_n) = \Psi(w_n)\Psi(w_1 \dots w_{n-1}) = \Psi(w_n)\Psi(w_1) \dots \Psi(w_{n-1}).$$

The first case is again in correspondence with the fact that Ψ is a morphism. The only cases which remain are for $n = 2$

$$\Psi(w_1)\Psi(w_2) = \Psi(w_2)\Psi(w_1)$$

and for $n > 2$

$$\Psi(w) = \Psi(w_2) \dots \Psi(w_n)\Psi(w_1) = \Psi(w_n)\Psi(w_1) \dots \Psi(w_{n-1}).$$

In both cases, since Ψ is a bijection, we get $w_1 = w_2 = \dots = w_n$. Hence, again $\Psi(w) = \Psi(w_1)\Psi(w_2) \dots \Psi(w_n)$.

With the same reasoning, we deduce that if $\Psi(w) = \Psi(w_{n-1}) \dots \Psi(w_2)\Psi(w_1)$ for every $w = w_1w_2 \dots w_{n-1} \in \mathcal{A}^*$, $n \geq 2$, then for an arbitrary $w = w_1w_2 \dots w_n \in \mathcal{A}^*$, $w_i \in \mathcal{A}$, we get $\Psi(w) = \Psi(w_n) \dots \Psi(w_2)\Psi(w_1)$. \square

Observation 18. Let \mathbf{u} be an infinite word whose language is invariant under a symmetry Ψ . For every w in $\mathcal{L}(\mathbf{u})$ whose frequency exists, $\rho(w) = \rho(\Psi(w))$ holds.

Remark 19. If a finite set G is a submonoid of $AM(\mathcal{A}^*)$, then G is a group and any of its members restricted to the set of words of length one is just a permutation on the alphabet \mathcal{A} . In other words, G is a finite group of symmetries. Words with languages invariant under all elements of such a group G of symmetries were studied in [13].

4. Factor Frequencies of Languages Invariant Under More Symmetries

Assume \mathbf{u} is an infinite word over an alphabet \mathcal{A} with $\#\mathcal{A} \geq 2$ whose language is invariant under all elements of a finite group $G \subset AM(\mathcal{A}^*)$ of symmetries containing an antimorphism. Let us summarize some observations concerning the group G of symmetries and reduced Rauzy graphs of \mathbf{u} . These observations constitute all tools we need for the proof of the main theorem of this paper - Theorem 20.

Observations:

1. Let θ be an antimorphism in G . The mapping $\Psi \rightarrow \theta\Psi$ is a bijection on G satisfying

$$\Psi \in G \text{ is a morphism} \iff \theta\Psi \in G \text{ is an antimorphism.}$$

This implies that G containing an antimorphism has an even number of elements, i.e., $\#G = 2k$.

2. For a factor w containing all letters of \mathcal{A} , the following properties can be easily verified:
 - (a) for any distinct antimorphisms $\theta_1, \theta_2 \in G$, we have $\theta_1(w) \neq \theta_2(w)$,
 - (b) for any distinct morphisms $\varphi_1, \varphi_2 \in G$, we have $\varphi_1(w) \neq \varphi_2(w)$.
3. If w is a θ -palindrome containing all letters of \mathcal{A} for an antimorphism $\theta \in G$, then θ is an involution, i.e., $\theta^2 = \text{Id}$.
4. In a reduced Rauzy graph of \mathbf{u} , if there is an edge e between two vertices w and v , where w and v contain all letters of \mathcal{A} , then
 - (a) either e is a θ -palindrome for some antimorphism $\theta \in G$, then there exist at least $\frac{\#G}{2}$ distinct edges having the same label $\rho(e)$, namely edges $\varphi(e)$ for all morphisms in G ;
 - (b) or e is not a θ -palindrome for any antimorphism $\theta \in G$, then there exist at least $\#G$ distinct edges having the same label $\rho(e)$, namely edges $\varphi(e)$ for all morphisms in G and $\theta(e)$ for all antimorphisms in G .
5. On one hand, if an edge e in the reduced Rauzy graph $\tilde{\Gamma}_n$ is mapped by an antimorphism θ onto itself, then the corresponding simple path has a θ -palindromic central factor of length n or $n + 1$. On the other hand, every θ -palindrome contained in $\mathcal{L}_{n+1}(\mathbf{u})$ is the central factor of a simple path mapped by θ onto itself and every θ -palindrome of length n is either the central factor of a simple path mapped by θ onto itself or is a special factor (thus, evidently, a BS factor).

Theorem 20. *Let $G \subset AM(\mathcal{A}^*)$ be a finite group containing an antimorphism and let \mathbf{u} be a uniformly recurrent aperiodic infinite word whose language is invariant under all elements of G and such that the frequency $\rho(w)$ exists for every factor $w \in \mathcal{L}(\mathbf{u})$. Then there exists $N \in \mathbb{N}$ such that*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \leq \frac{1}{\#G} \left(4\Delta\mathcal{C}(n) + \#G - X - Y \right) \quad \text{for all } n \geq N,$$

where X is the number of BS factors of length n and Y is the number of BS factors of length n that are θ -palindromes for an antimorphism $\theta \in G$.

Proof. Since \mathbf{u} is uniformly recurrent, we can find N such that any factor of length N contains all letters of \mathbf{u} . Let $\tilde{\Gamma}_n$ be the reduced Rauzy graph of \mathbf{u} of order $n \geq N$. We know already that the set of edge labels of $\tilde{\Gamma}_n$ is equal to the set of edge labels of Γ_n . It is easy to see that any element of G is an automorphism of $\tilde{\Gamma}_n$, i.e., G maps the graph $\tilde{\Gamma}_n$ onto itself.

Let A denote the number of edges e in $\tilde{\Gamma}_n$ such that e is mapped by a certain antimorphism of G onto itself (such an antimorphism is involutive by Observation (3)) and by B the number of edges e in $\tilde{\Gamma}_n$ such that e is not mapped by any antimorphism of G onto itself, then

$$\#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\} = A + B \leq 3\Delta\mathcal{C}(n) - X, \tag{10}$$

where the upper bound is taken from (6). We get, using Observations (3) and (5), the following formula

$$A = \sum_{\theta \in G^{(2)}} (\mathcal{P}_\theta(n) + \mathcal{P}_\theta(n+1)) - \sum_{\theta \in G^{(2)}} \#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w = \theta(w) \text{ and } w \text{ BS}\}, \tag{11}$$

where we subtract the number of BS factors of $\mathcal{L}_n(\mathbf{u})$ that are θ -palindromes for a certain antimorphism θ , in the statement denoted by Y , since they are not central factors of any simple path. If $\#G = 2k$, then for every edge e in $\tilde{\Gamma}_n$ that is mapped by a certain antimorphism $\theta \in G$ onto itself, there are at least k different edges with the same label $\rho(e)$ by Observation (4a).

Now, let us turn our attention to those edges of $\tilde{\Gamma}_n$ which are not mapped by any antimorphism of G onto themselves. For every such edge e , at least $2k$ edges have the same label $\rho(e)$ by Observation (4b). These considerations lead to the following estimate

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \leq \frac{1}{k}A + \frac{1}{2k}B = \frac{1}{2k}A + \frac{1}{2k}(A + B). \tag{12}$$

Putting together (11), (10), (12), and Theorem 2, the statement is proven. □

Remark 21. If an infinite word \mathbf{u} is closed under reversal, then $G = \{\text{Id}, R\}$ and the new upper bound from Theorem 20 coincides with the estimate from Theorem 12.

Finally, if we want to have a simpler upper bound on factor frequencies, we can use the following one, which is slightly rougher than the estimate from Theorem 20.

Corollary 22. *Let $G \subset AM(\mathcal{A}^*)$ be a finite group containing an antimorphism and let \mathbf{u} be a uniformly recurrent infinite word whose language is invariant under all elements of G and such that the frequency $\rho(w)$ exists for every factor $w \in \mathcal{L}(\mathbf{u})$. Then there exists $N \in \mathbb{N}$ such that*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \leq \frac{4}{\#G} \Delta C(n) + 1 \quad \text{for all } n \geq N.$$

The equality holds for all sufficiently large n if and only if \mathbf{u} is purely periodic.

5. Optimality of the Upper Bound

In this section, we will illustrate on an example taken from [13] that the upper bound from Theorem 20 is attained for every $n \in \mathbb{N}$, $n \geq 1$, thus it is an optimal upper bound. The infinite word \mathbf{u} in question is the fixed point starting in 0, which is obtained when we iterate the primitive morphism φ given by:

$$\varphi(0) = 0130, \quad \varphi(1) = 1021, \quad \varphi(2) = 102, \quad \varphi(3) = 013, \quad (13)$$

i.e., for all $n \in \mathbb{N}$, the word $\varphi^n(0)$ is a prefix of \mathbf{u} .

The corresponding incidence matrix is of the form

$$M_\varphi = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

its dominant eigenvalue is $\lambda = 2 + \sqrt{3}$ with the corresponding normalized eigenvector

$$\frac{1}{2} \begin{pmatrix} \sqrt{3} - 1 \\ \sqrt{3} - 1 \\ 2 - \sqrt{3} \\ 2 - \sqrt{3} \end{pmatrix},$$

hence we get the letter frequencies

$$\rho(0) = \rho(1) = \frac{\sqrt{3} - 1}{2}, \quad \rho(2) = \rho(3) = \frac{2 - \sqrt{3}}{2}.$$

We also know that the frequencies of all factors exist because of the primitivity of φ . In [13], the following properties of \mathbf{u} were shown:

1. The language $\mathcal{L}(\mathbf{u})$ is closed under the finite group of symmetries $G = \{\text{Id}, \theta_1, \theta_2, \theta_1\theta_2\}$, where θ_1, θ_2 are involutive antimorphisms acting on \mathcal{A} as follows:

$$\theta_1 : 0 \rightarrow 1, 1 \rightarrow 0, 2 \rightarrow 2, 3 \rightarrow 3 \quad \text{and} \quad \theta_2 : 0 \rightarrow 0, 1 \rightarrow 1, 2 \rightarrow 3, 3 \rightarrow 2.$$
2. The first increment of factor complexity satisfies $\Delta\mathcal{C}(n) = 2$ for all $n \in \mathbb{N}$, $n \geq 1$. Moreover, every LS factor w is a prefix for some $n \in \mathbb{N}$,
 - either of $\varphi^n(0) = 013010210130130\dots$ and $\text{Lext}(w) = \{1, 3\}$
 - or of $\varphi^n(1) = 102101301021021\dots$ and $\text{Lext}(w) = \{0, 2\}$.
3. A factor w of \mathbf{u} is LS if and only if $\theta_i(w)$ is RS for $i \in \{1, 2\}$.

In order to find the set of frequencies of factors of any length, we need to describe BS factors of \mathbf{u} . By Property (3), we deduce the relation between BS factors and θ_i -palindromes.

Corollary 23. *Every nonempty BS factor is a θ_i -palindrome for one of the indices $i \in \{1, 2\}$.*

Proposition 24. *If $v \in \mathcal{L}(\mathbf{u})$ is a BS factor of length greater than 5, then $v = \varphi(w)p_{w_n}$, where w_n is the last letter of w and $p_0 = p_2 = 10210$ and $p_1 = p_3 = 01301$. Moreover,*

$$\rho(v) = \frac{1}{\lambda} \rho(w).$$

Proof. By Property (2), every LS factor of length greater than 5 starts either in 01301 or in 10210. Similarly, by Property (3), every RS factor ends either in 01301 or in 10210. It follows from the definition of φ in (13) that there exists $w \in \mathcal{L}(\mathbf{u})$ such that $v = \varphi(w)01301$ or $v = \varphi(w)10210$ and that w is necessarily a BS factor. Consider $v = \varphi(w)01301$, the second case can be treated analogously. It is then not difficult to see that w ends in $w_n = 1$ or $w_n = 3$, hence $p_{w_n} = 01301$. In order to prove the relation between frequencies, we need to determine the set of interpretations of v . It is readily seen that the set of interpretations is

- $\{(w01, 0, 3), (w02, 0, 2), (w30, 0, 2)\}$ if $w_n = 1$ or $w_n = 3$,
- $\{(w10, 0, 3), (w13, 0, 2), (w21, 0, 2)\}$ if $w_n = 0$ or $w_n = 2$.

Using Proposition 5, we obtain $\rho(v) = \frac{\rho(w01)+\rho(w02)+\rho(w30)}{\lambda} = \frac{\rho(w)}{\lambda}$ if $w_n = 1$ or $w_n = 3$, where the last equality follows from the fact that w is always followed by 01, 02, or 30, and similarly, $\rho(v) = \frac{\rho(w10)+\rho(w13)+\rho(w21)}{\lambda} = \frac{\rho(w)}{\lambda}$ if $w_n = 0$ or $w_n = 2$. □

Proposition 24 implies that if we want to generate all BS factors of \mathbf{u} , then it is enough to know non-empty BS factors of length less than or equal to 5 and to apply the mapping $w \rightarrow \varphi(w)p_{w_n}$ on them repeatedly. Nonempty BS factors of length less than or equal to 5 are:

1. 0 and 1,
2. 01 and 10,
3. 01301 and 10210.

The aim of the rest of this section is to show that for any length $n \in \mathbb{N}$, $n \geq 1$, we have

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} = \begin{cases} 2 & \text{if } \mathcal{L}_n(\mathbf{u}) \text{ contains a BS factor,} \\ 3 & \text{otherwise.} \end{cases}$$

Let us draw in Figure 1 reduced Rauzy graphs containing short BS factors. In

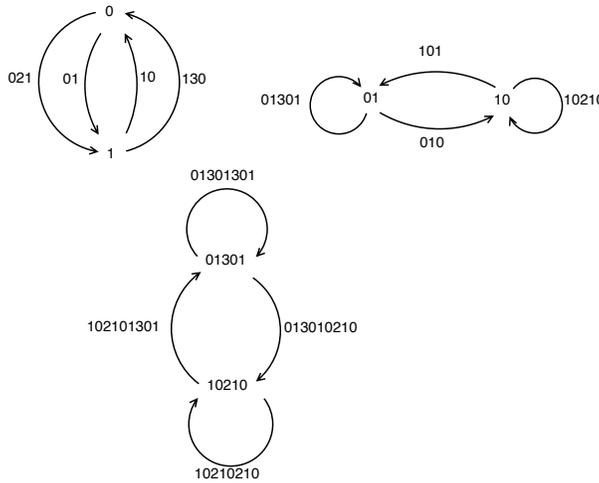


Figure 1: Reduced Rauzy graphs of \mathbf{u} of order $n \in \{1, 2, 5\}$.

order to describe factor frequencies, it suffices to consider reduced Rauzy graphs containing short BS factors together with the following observations concerning reduced Rauzy graphs of \mathbf{u} .

- Observation 25.**
1. Any reduced Rauzy graph of \mathbf{u} has either four vertices (two LS factors and two RS factors) or two vertices (BS factors).
 2. Reduced Rauzy graphs of order larger than 5 whose vertices are BS factors are obtained from the graphs in Figure 1 by a repeated application of the mapping $w \rightarrow \varphi(w)p_{w_n}$ simultaneously to all vertices and edges.
 3. By Corollary 8, it is not difficult to see that if we find to a reduced Rauzy graph $\tilde{\Gamma}_n$ whose vertices are not BS factors the reduced Rauzy graph of minimal larger order, say $\tilde{\Gamma}_m$, whose vertices are BS factors, then

$$\{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_n\} = \{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_m\} \cup \{\rho(v) \mid v \text{ vertex in } \tilde{\Gamma}_m\}.$$

The last step in the derivation of frequencies of factors of \mathbf{u} is to determine the frequencies of edges and vertices in the reduced Rauzy graphs depicted in Figure 1. In the sequel, we make use of the Kirchhoff law for frequencies (Observation 6), of the fact that symmetries preserve factor frequencies, and of the formula from Proposition 5.

1. $\tilde{\Gamma}_1$:

$$\begin{aligned} \rho(0) = \rho(1) &= \frac{\sqrt{3}-1}{2} = \frac{\sqrt{3}+1}{2\lambda} \\ \rho(130) = \rho(021) &= \rho(2) = \frac{2-\sqrt{3}}{2} = \frac{1}{2\lambda} \\ \rho(01) = \rho(10) &= \rho(0) - \rho(130) = \frac{\sqrt{3}}{2\lambda} \end{aligned}$$

In the second row, the first equality follows from the fact that symmetries preserve frequencies and $130 = \theta_2(021)$ and the second equality by Corollary 7 from the fact that 2 is neither LS, nor RS. In the third row, the first equality is again due to symmetries and the second uses the Kirchhoff law for frequencies from Observation 6.

2. $\tilde{\Gamma}_2$:

$$\begin{aligned} \rho(01) = \rho(10) &= \frac{\sqrt{3}}{2\lambda} \\ \rho(01301) = \rho(10210) &= \rho(130) = \frac{1}{2\lambda} \\ \rho(010) = \rho(101) &= \rho(01) - \rho(01301) = \frac{\sqrt{3}-1}{2\lambda} \end{aligned}$$

3. $\tilde{\Gamma}_5$:

$$\begin{aligned} \rho(01301) = \rho(10210) &= \frac{1}{2\lambda} \\ \rho(\varphi(0)10210) = \rho(\varphi(1)01301) &= \frac{\rho(0)}{\lambda} = \frac{\sqrt{3}-1}{2\lambda} \\ \rho(01301301) = \rho(10210210) &= \rho(01301) - \rho(\varphi(0)10210) = \frac{2-\sqrt{3}}{2\lambda} = \frac{1}{2\lambda^2} \end{aligned}$$

Putting together Proposition 24, properties of reduced Rauzy graphs summarized in Observation 25, and the knowledge of frequencies of vertices and edges in $\tilde{\Gamma}_1, \tilde{\Gamma}_2,$ and $\tilde{\Gamma}_5$, we obtain the following corollary.

Corollary 26. *Let $n \in \mathbb{N}, n \geq 1$, such that*

1. $\mathcal{L}_n(\mathbf{u})$ contains a BS factor: then there exists $k \in \mathbb{N}$ such that the set $\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\}$ is of one of the following forms:
 - (a) $\{\frac{1}{2\lambda^{k+1}}, \frac{\sqrt{3}}{2\lambda^{k+1}}\},$
 - (b) $\{\frac{1}{2\lambda^{k+1}}, \frac{\sqrt{3}-1}{2\lambda^{k+1}}\},$
 - (c) $\{\frac{\sqrt{3}-1}{2\lambda^{k+1}}, \frac{1}{2\lambda^{k+2}}\}.$
2. $\mathcal{L}_n(\mathbf{u})$ does not contain a BS factor: then there exists $k \in \mathbb{N}$ such that the set $\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\}$ is of one of the following forms:

- (a) $\{\frac{\sqrt{3}-1}{2\lambda^k}, \frac{1}{2\lambda^{k+1}}, \frac{\sqrt{3}}{2\lambda^{k+1}}\},$
- (b) $\{\frac{\sqrt{3}}{2\lambda^{k+1}}, \frac{1}{2\lambda^{k+1}}, \frac{\sqrt{3}-1}{2\lambda^{k+1}}\},$
- (c) $\{\frac{1}{2\lambda^{k+1}}, \frac{\sqrt{3}-1}{2\lambda^{k+1}}, \frac{1}{2\lambda^{k+2}}\}.$

A direct consequence of the previous corollary is the optimality of the upper bound from Theorem 20.

Proposition 27. *Let \mathbf{u} be the fixed point of φ defined in (13). Then for every $n \in \mathbb{N}, n \geq 1$, we have*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} = \frac{1}{\#G} (4\Delta C(n) + \#G - X - Y),$$

where X is the number of BS factors of length n and Y is the number of BS factors of length n that are θ_1 - or θ_2 -palindromes.

Proof. Let us consider at first n such that $\mathcal{L}_n(\mathbf{u})$ does not contain a BS factor. Then, on one hand, Corollary 26 states that $\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} = 3$. On the other hand, $\frac{1}{\#G} (4\Delta C(n) + \#G - X - Y) = \frac{4 \cdot 2 + 4 - 0 - 0}{4} = 3$. At second, let $\mathcal{L}_n(\mathbf{u})$ contain a BS factor. Then, on one hand, we have by Corollary 26 that $\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} = 2$. On the other hand, by (1) of Observation 25, $\mathcal{L}_n(\mathbf{u})$ contains 2 BS factors, and by Corollary 23, one BS factor is a θ_1 -palindrome and one BS factor is a θ_2 -palindrome, thus $\frac{1}{\#G} (4\Delta C(n) + \#G - X - Y) = \frac{4 \cdot 2 + 4 - 2 - 2}{4} = 2$. \square

Remark 28. There are also infinite words whose language is invariant under elements of a finite group of symmetries, however, the upper bound from Theorem 20 is not reached for any $n \geq N$. Such an example is the famous Thue-Morse word. Its group of symmetries $G = \{\text{Id}, R, \Psi, \Psi \circ R\}$, where Ψ is a morphism acting on $\{0, 1\}$ as follows:

$$\Psi : 0 \rightarrow 1, 1 \rightarrow 0.$$

As shown by Dekking [8], the Thue-Morse word \mathbf{u}_{TM} satisfies for $n \in \mathbb{N}, n \geq 1$,

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u}_{TM})\} = \begin{cases} 1 \text{ or } 2 & \text{if } \mathbf{u}_{TM} \text{ contains a BS factor of length } n, \\ 2 & \text{otherwise.} \end{cases}$$

But, the upper bound from Theorem 20 is of the following form for $n \in \mathbb{N}, n \geq 1$,

$$\begin{cases} 2 \text{ or } 4 & \text{if } \mathbf{u}_{TM} \text{ contains a BS factor of length } n, \\ 3 \text{ or } 5 & \text{otherwise.} \end{cases}$$

Acknowledgement I would like to thank E. Pelantová and Š. Starosta for careful reviewing and useful remarks. I acknowledge financial support by the Czech Science Foundation grant 201/09/0584, by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic.

References

- [1] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, *Palindrome complexity*, Theoret. Comput. Sci. **292** (2003), 9–31
- [2] L. Balková, E. Pelantová, *A note on symmetries in the Rauzy graph and factor frequencies*, Theoret. Comput. Sci. **410** (2009), 2779–2783
- [3] P. Baláži, Z. Masáková, E. Pelantová, *Factor versus palindromic complexity of uniformly recurrent infinite words*, Theoret. Comput. Sci. **380** (2007), 266–275
- [4] V. Berthé, *Fréquences des facteurs des suites sturmiennes*, Theoret. Comput. Sci. **165** (1996), 295–309
- [5] M. Boshernitzan, *A condition for unique ergodicity of minimal symbolic flows*, Ergodic Theory Dynam. Systems **12** (1992), 425–428
- [6] J. Cassaigne, *Special factors of sequences with linear subword complexity*, Developments in language theory, II (Magdeburg, 1995), World Sci. Publishing, Singapore (1996), 25–34
- [7] J. Cassaigne, *Complexité et facteurs spéciaux [Complexity and special factors]*, Journées Montoises (Mons, 1994), Bull. Belg. Math. Soc. Simon Stevin **4** (1997), 67–88
- [8] M. Dekking, *On the Thue-Morse measure*, Acta Univ. Carolin. Math. Phys. **33** (1992), 35–40
- [9] S. Ferenczi, L. Zamboni, *Languages of k -interval exchange transformations*, Bull. Lond. Math. Soc. **40** (2008), 705–714
- [10] A. Frid, *On the frequency of factors in a DOL word*, J. Autom. Lang. Comb. **3** (1998), 29–41
- [11] A. Glen, J. Justin, S. Widmer, L. Q. Zamboni, *Palindromic richness*, European J. Combin. **30** (2009), 510–531
- [12] E. Pelantová, Š. Starosta, *Infinite words rich and almost rich in generalized palindromes*, in: G. Mauri, A. Leporati (Eds.), Developments in Language Theory, volume 6795 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg (2011), 406–416
- [13] E. Pelantová, Š. Starosta, *Languages invariant under more symmetries: overlapping factors versus palindromic richness*, arXiv:1103.4051 [math.CO]
- [14] E. Pelantová, Š. Starosta, *Palindromic richness and Coxeter groups*, arXiv:1108.3042 [math.CO]
- [15] M. Queffélec, *Substitution dynamical systems - Spectral analysis*, in Lecture Notes in Math. **1294**, 1987
- [16] Š. Starosta, *On theta-palindromic richness*, Theoret. Comput. Sci. **412** (2011), 1111–1121
- [17] N. Wozny, L. Q. Zamboni, *Frequencies of factors in Arnoux-Rauzy sequences*, Acta Arith. **96** **3** (2001), 261–278