KYBERNETIKA — VOLUME *48* (2012), NUMBER 3, PAGES 371–385

FACTOR FREQUENCIES IN GENERALIZED THUE–MORSE WORDS

Ľubomíra Balková

We describe factor frequencies of the generalized Thue–Morse word $\mathbf{t}_{b,m}$ defined for $b \geq 2$, $m \geq 1$, $b, m \in \mathbb{N}$, as the fixed point starting in 0 of the morphism

$$\varphi_{b,m}(k) = k(k+1)\dots(k+b-1),$$

where $k \in \{0, 1, \dots, m-1\}$ and where the letters are expressed modulo m. We use the result of Frid [4] and the study of generalized Thue–Morse words by Starosta [6].

Keywords: combinatorics on words, generalized Thue–Morse word, factor frequency

Classification: 68R15

1. INTRODUCTION

The generalized Thue–Morse word $\mathbf{t}_{b,m}$ is defined for $b \geq 2$, $m \geq 1$, $b, m \in \mathbb{N}$, as the fixed point starting in 0 of the morphism

$$\varphi_{b,m}(k) = k(k+1)\dots(k+b-1),$$

where $k \in \{0, 1, \dots, m-1\}$ and where the letters are expressed modulo m. Naturally, the class of generalized Thue–Morse words contains the famous Thue–Morse word $\mathbf{t}_{2,2}$ whose factor frequencies have been determined by Dekking [3].

Generalized Thue–Morse words belong to the class of circular fixed points of uniform marked primitive morphisms. For such a class, Frid [4] has provided an algorithm for the computation of factor frequencies. We recall her algorithm in Section 1. The aim of this paper is to describe the set of frequencies of factors of length n in $\mathbf{t}_{b,m}$ for every $n \in \mathbb{N}$. The most direct way is to apply Frid's algorithm. However, there is even an easier way thanks to the knowledge of reduced Rauzy graphs (obtained from the description of bispecial factors by Starosta [6]) and the invariance of the generalized Thue–Morse word under symmetries preserving factor frequencies. In Section 2, we define reduced Rauzy graphs and their relation to factor frequencies. In Section 3, we explain what a symmetry is and how it preserves factor frequencies. The main result is presented in Section 4, where we combine Frid's algorithm, reduced Rauzy graphs, and symmetries in order to get factor frequencies of generalized Thue–Morse words. Recently, an optimal upper bound on the number of factor frequencies in infinite words whose language is invariant under more symmetries has been derived in [2]. The generalized Thue–Morse word is an example of infinite words for which the upper bound is not attained, as shown in Section 5.

We ask the reader to consult Preliminaries of the paper Generalized Thue–Morse words and palindromic richness by Starosta [6] for undefined terms.

2. FACTOR FREQUENCIES OF FIXED POINTS OF CIRCULAR MARKED UNIFORM MORPHISMS

If w is a factor of an infinite word $\mathbf{u}$ and if the following limit exists

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

where $\mathcal{L}(\mathbf{u})$ denotes the set of factors of $\mathbf{u}$, then the limit is denoted $\rho(w)$ and called the *frequency* of w in $\mathbf{u}$. Generalized Thue–Morse words are fixed points of primitive morphisms, therefore, in the sequel, we limit our considerations to primitive morphisms.

Let us recall first a result of Frid [4], 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 $\mathcal{A}^* = \{a_1, a_2, \ldots, a_m\}^*$. 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 [5], for fixed points of primitive morphisms,

- factor frequencies exist,
- it follows from the Perron–Frobenius theorem that the incidence matrix has one real eigenvalue λ which is larger than the modulus of any other eigenvalue,
- the components of the unique eigenvector $(x_1, x_2, \ldots, x_m)^T$ 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 $i \in \{1, 2, \ldots, m\}$.

Let φ be a morphism on $\mathcal{A}^*$. We denote $\psi_{ij} : \mathcal{A}^+ \to \mathcal{A}^+$, where $i, j \in \mathbb{N}$, the mapping that associates to $v \in \mathcal{A}^+$ 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* $s = (b_0 b_1 \ldots b_m, i, j)$ if $v = \psi_{ij}(b_0 b_1 \ldots b_m)$, where $b_k \in \mathcal{A}$ and $i < |\varphi(b_0)|$ and $j < |\varphi(b_m)|$. The word $a(s) = b_0 b_1 \ldots b_m$ is an *ancestor* of s. The set of all interpretations of v is denoted $I(v)$. Now, we can recall the result of Frid for factor frequencies of fixed points of primitive morphisms.

Proposition 2.1. 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 of $\mathbf{u}$, it holds

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

For circular fixed points of uniform marked primitive morphisms, the algorithm of Frid [4] provides the possible frequencies of factors of a given length and for every frequency, the number of factors having that frequency. In order to describe her algorithm, we have to recall several notions. A morphism φ defined on the alphabet $\mathcal{A}$ is called *uniform* if all images of letters are of the same length b, i. e., $|\varphi(a)| = b$ for all $a \in \mathcal{A}$. In the case of a uniform primitive morphism φ, the dominant eigenvalue of the incidence matrix M_φ is $\lambda = b$. A morphism is called *marked* (sometimes also *bifix-free*) if every pair of images of distinct letters differs both in the first letter and in the last letter. Let $\mathbf{u}$ be a fixed point of a morphism φ defined on $\mathcal{A}$, then its factor w contains a *synchronization point* (w_1, w_2) if $w = w_1 w_2$ and for every $v_1, v_2 \in \mathcal{A}^*$ and for every factor s of $\mathbf{u}$, there exists factors s_1, s_2 of $\mathbf{u}$ such that the following implication holds

$$v_1 w v_2 = \varphi(s) \Rightarrow s = s_1 s_2,\ v_1 w_1 = \varphi(s_1),\ w_2 v_2 = \varphi(s_2).$$

In other words, a synchronization point marks a boundary between letter images in every occurrence of w in $\mathbf{u}$. Any factor w of $\mathbf{u}$ that contains a synchronization point is called *circular*. We call a fixed point $\mathbf{u}$ of a morphism φ *circular (with synchronization delay L)* if every factor w of length greater than or equal to L is circular. For uniform marked primitive morphisms, Proposition 2.1 takes the following form (as provided in [4]).

Proposition 2.2. Let v be a circular factor of a fixed point of a uniform marked primitive morphism φ with the letter image length b, then there exists a unique interpretation of v. Moreover, if we denote the unique ancestor of v by w, then $\rho(v) = \frac{\rho(w)}{b}$.

We define the *structure ordering number* K for fixed points of circular uniform morphisms as the least integer satisfying $b(K-1)+1 \geq L$. The following statements can be found in [4] as Proposition 4 and Theorem 5.

Proposition 2.3. Let $n \geq K$, then there exists a unique triplet of decomposition parameters $(p(n), k(n), \Delta(n))$, where $p(n) \in \mathbb{N}$, $k(n) \in \{K, \ldots, b(K-1)\}$, and $\Delta(n) \in \{1, \ldots, b^{p(n)}\}$, such that

$$n = b^{p(n)}(k(n) - 1) + \Delta(n).$$

The explicit formulae read $p(n) = \left\lceil \log_b \frac{n}{K-1} \right\rceil - 1$, $k(n) = \left\lceil \frac{n}{b^{p(n)}} \right\rceil$, $\Delta(n) = n - b^{p(n)}(k(n) - 1)$. Let us recall that $\mathcal{L}_n(\mathbf{u})$ denotes the set of factors of $\mathbf{u}$ of length n and $\mathcal{C}$ denotes the factor complexity.

Theorem 2.4. Let $\mathbf{u}$ be a circular fixed point of a uniform marked primitive morphism φ. Denote $\mathcal{L}_n(\mathbf{u}) = \{v_1^{(n)}, v_2^{(n)}, \ldots, v_{\mathcal{C}(n)}^{(n)}\}$, where the index n emphasizes the length. For all $n \geq K$, the set $\mathcal{L}_{n+1}(\mathbf{u})$ can be partitioned into

(a) $\mathcal{C}(k(n)+1)$ groups of $\Delta(n)$ words each, every word in the jth group having the frequency $\frac{1}{b^{p(n)}} \rho\big(v_j^{k(n)+1}\big)$, $j \in \{1, \ldots, \mathcal{C}(k(n)+1)\}$,

(b) $\mathcal{C}(k(n))$ groups of $b^{p(n)} - \Delta(n)$ words each, every word in the jth group having the frequency $\frac{1}{b^{p(n)}} \rho\big(v_j^{k(n)}\big)$, $j \in \{1, \ldots, \mathcal{C}(k(n))\}$.

The frequencies $\rho\big(v_j^{(k)}\big)$, $k \in \{K, \ldots, b(K-1)+1\}$, can be found directly using (1). Theorem 2.4 provides then explicit formulae for factor frequencies of circular fixed points of uniform marked primitive morphisms.

2.1. 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 $\rho(e)$.

It is easy to see that edge frequencies in a Rauzy graph Γ_n behave similarly as the current in a circuit. We may formulate an analogy of the Kirchhoff's current law: the sum of frequencies of edges ending in a vertex equals the sum of frequencies of edges starting in this vertex.

Observation 2.5. (Kirchhoff's law for frequencies) Let w be a factor of an infinite word $\mathbf{u}$ over the alphabet $\mathcal{A}$ whose factor frequencies exist. Then

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

where $\mathrm{Lext}(w)$ denotes the set of left extensions of w, i. e., $\mathrm{Lext}(w) = \{a \in \mathcal{A} \mid aw \in \mathcal{L}(\mathbf{u})\}$ and $\mathrm{Rext}(w)$ denotes the set of right extensions.

The Kirchhoff's law for frequencies has some useful consequences.

Corollary 2.6. Let w be a factor of an infinite word $\mathbf{u}$ such that $\rho(w)$ 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 2.7. Let w be a factor of an aperiodic recurrent infinite word $\mathbf{u}$ such that $\rho(w)$ exists. Let v be the shortest bispecial (BS) factor containing w, then $\rho(w) = \rho(v)$.

The assumption of recurrence and aperiodicity in Corollary 2.7 is needed in order to ensure that every factor can be extended to a BS factor.

Corollary 2.6 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 outline this construction, we introduce the notion of an n-simple path.

Definition 2.8. Let $\mathbf{u}$ be an infinite word whose factor frequencies exist. A factor e of length larger than n such that its prefix and its suffix of length n are special factors of $\mathbf{u}$ and e does not contain any other special factors is called an n-simple path. We define the label of an n-simple path e as $\rho(e)$.

Definition 2.9. The reduced Rauzy graph $\tilde{\Gamma}_n$ of $\mathbf{u}$ of order n is a directed graph whose set of vertices is formed by left special (LS) and right special (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 an n simple path starting in w and ending in v. We assign to such an edge e the label of the corresponding n-simple path.

Remark 2.10. According to Corollary 2.6 and Definition 2.9, if $\mathbf{u}$ is an aperiodic recurrent infinite word whose factor frequencies exist, it holds for every $n \in \mathbb{N}$ that

$$\{\rho(e) \bigm| e \in \mathcal{L}_{n+1}(\mathbf{u})\} = \{\rho(e) \bigm| e \text{ edge in } \tilde{\Gamma}_n\}.$$

Considering Corollary 2.7 and Definition 2.9, one may observe the following.

Observation 2.11. Let $\mathbf{u}$ be an aperiodic recurrent infinite word whose factor frequencies exist. Take $n \in \mathbb{N}$ such that $\mathbf{u}$ does not contain a BS factor of length n and denote by m the minimal number $m > n$ such that $\mathcal{L}_m(\mathbf{u})$ contains a BS factor. Then

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

Similarly as in the above observation, we usually say a BS vertex instead of a vertex of the reduced Rauzy graph $\tilde{\Gamma}_n$ corresponding to a BS factor in $\mathcal{L}_n(\mathbf{u})$.

3. SYMMETRIES PRESERVING FACTOR FREQUENCY

We will be interested in symmetries preserving in a certain way the number of 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: $\mathcal{A}^* \to \mathcal{A}^*$,
2. for all $w, v \in \mathcal{A}^*$

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

The following statements are taken from [2]. Recall that $\theta : \mathcal{A}^* \to \mathcal{A}^*$ is an antimorphism if for any $w, v \in \mathcal{A}^*$ it satisfies $\theta(wv) = \theta(v)\theta(w)$ and w is a θ-palindrome if $\theta(w) = w$.

Theorem 3.1. Let $\Psi : \mathcal{A}^* \to \mathcal{A}^*$. Then Ψ is a symmetry if and only if Ψ is a morphism or an antimorphism such that Ψ restricted to $\mathcal{A}$ is a letter permutation.

Observation 3.2. 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, the frequency $\rho(\Psi(w))$ exists as well and it holds

$$\rho(w) = \rho(\Psi(w)).$$

We denote the set of all morphisms and antimorphisms on $\mathcal{A}^*$ by $AM(\mathcal{A}^*)$.

Theorem 3.3. 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 $M \in \mathbb{N}$ such that

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \le \frac{1}{\#G}\Big(4\big(\mathcal{C}(n+1) - \mathcal{C}(n)\big) + \#G - X - Y\Big) \quad \text{for all } n \ge M,$$

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 $\theta \in G$.

4. FACTOR FREQUENCIES OF GENERALIZED THUE–MORSE WORDS

The generalized Thue–Morse word $\mathbf{t}_{b,m}$ is defined for $b \geq 2,\ m \geq 1,\ b, m \in \mathbb{N}$, as the fixed point starting in 0 of the morphism

$$\varphi_{b,m} : k \ \rightarrow \ k(k+1)\dots(k+b-1), \tag{2}$$

where $k \in \mathbb{Z}_m = \{0, 1, \dots, m-1\}$ and where the letters are expressed modulo m.

Definition 4.1. We denote by q the smallest positive integer satisfying $q(b-1) \equiv 0 \mod m$.

The word $\mathbf{t}_{b,m}$ is periodic if and only if $b \equiv 1 \mod m$ (see [1]). In this case, $\mathbf{t}_{b,m} = (01\dots(m-1))^\omega$, where ω denotes an infinite repetition. It is thus readily seen that any factor of $\mathbf{t}_{b,m}$ has its frequency equal to $\frac{1}{m}$.

Properties of $\varphi_{b,m}$ and $\mathbf{t}_{b,m}$:

a) $\varphi_{b,m}$ is primitive, thus letter frequencies exist and are equal to the components of the eigenvector $\frac{1}{m}(1, 1, \dots, 1)^T$ of the incidence matrix corresponding to the dominant eigenvalue b,

b) $\varphi_{b,m}$ is uniform ($|\varphi_{b,m}(k)| = b$ for all $k \in \mathbb{Z}_m$),

c) $\varphi_{b,m}$ is marked,

d) $\mathbf{t}_{b,m}$ is circular with synchronization delay $L = 2b$.

Proof. Any $w \in \mathcal{L}(\mathbf{t}_{b,m})$ of length greater than or equal to $2b$ contains either for some $k, \ell \in \mathbb{Z}_m$ a factor $k\ell$, where $\ell \not\equiv k+1 \mod m$, or is of length $2b$ and of the form $w = k(k+1)\dots(k+2b-1)$ for some $k \in \mathbb{Z}_m$.

(i) In the first case, it is easy to see that k marks the end of w_1 and ℓ the beginning of w_2 in the synchronization point (w_1, w_2) of w.

(ii) In the second case, (w, ε) is a synchronization point of $w = k(k+1)\dots(k+2b-1)$, where ε denotes the empty word. □

e) $\mathcal{L}(\mathbf{t}_{b,m})$ is invariant under the dihedral group $D_m = \{\Pi_x \mid x \in \mathbb{Z}_m\} \cup \{\Psi_x \mid x \in \mathbb{Z}_m\}$, where for all $x \in \mathbb{Z}_m$, Π_x is a morphism and Ψ_x an antimorphism defined for all $k \in \mathbb{Z}_m$ by

$$\begin{aligned} \Pi_x(k) &= x+k, \\ \Psi_x(k) &= x-k. \end{aligned}$$

For the proof see [6].

In the sequel, we will only consider aperiodic words, i. e., $b \not\equiv 1 \mod m$. The aim of this section is to describe $\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{t}_{b,m})\}$ for all $n \in \mathbb{N}$. Theorem 2.4 gives explicit formulae for factor frequencies if the frequencies of factors of length $n \in \{1, \dots, 2b+1\}$ (the structure ordering number for $\mathbf{t}_{b,m}$ is $K = 3$) are known. There is even an easier way to get factor frequencies using symmetries of $\mathcal{L}(\mathbf{t}_{b,m})$ and the description of BS factors from [6].

Proposition 4.2. If v is a BS factor of $\mathcal{L}(\mathbf{t}_{b,m})$ of length greater than or equal to $2b$, then there exists a BS factor $w \in \mathcal{L}(\mathbf{t}_{b,m})$ such that $v = \varphi_{b,m}(w)$. Moreover, $\rho(v) = \frac{\rho(w)}{b}$.

Proof. The first part has been proved as Lemma 3 in [6]. The second part follows from Proposition 2.2. □

Remark 4.3. With Proposition 4.2 in hand, one can see that all BS factors are iterated images of BS factors of length less or equal to $2b-1$.

The set of BS factors of length n, where $1 \leq n \leq 2b-1$, taken from [6] reads:

$$\{w^{(j)} = j(j+1)\dots(j+n-1) \mid j \in \mathbb{Z}_m\}.$$

The corresponding set of left extensions is of the following form:

1. for $1 \leq n \leq b$
$$\mathrm{Lext}(w^{(j)}) = \{j-1+k(b-1) \mid k \in \{0,1,\dots,q-1\}\},$$

2. for $b+1 \leq n \leq 2b-1$
$$\mathrm{Lext}(w^{(j)}) = \{j-1, j+b-2\}.$$

There are no LS factors of length n which are not BS, where $1 \leq n \leq b$ and the set of LS factors of length n which are not BS, where $b+1 \leq n \leq 2b-1$, reads:

$$\{v^{(j)} = j(j+1)\dots(j+b-1)(j+1)\dots(j+n-b) \mid j \in \mathbb{Z}_m\}.$$

The corresponding set of left extensions is of the following form:

$$\mathrm{Lext}(v^{(j)}) = \{j-1+k(b-1) \mid k \in \{0,1,\dots,q-1\}\}.$$

Remark 4.4. It is not difficult to see that for any morphism Π of D_m, it holds:

1. w is a LS (RS) factor of $\mathbf{t}_{b,m}$ if and only if $\Pi(w)$ is a LS (RS) factor of $\mathbf{t}_{b,m}$,
2. left (right) extensions of $\Pi(w)$ are Π-images of left (right) extensions of w,

and for any antimorphism Ψ of D_m, it holds:

1. w is a LS (RS) factor of $\mathbf{t}_{b,m}$ if and only if $\Psi(w)$ is a RS (LS) factor of $\mathbf{t}_{b,m}$,
2. left (right) extensions of $\Psi(w)$ are Ψ-images of right (left) extensions of w.

Reduced Rauzy graph method (RRG method). We get the frequencies $\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{t}_{b,m})\}$ for all $n \in \mathbb{N}$ in the following way.

Step (i) We describe reduced Rauzy graphs of order n, in particular edge and vertex frequencies, where $1 \leq n \leq 2b-1$, making use of the invariance of $\mathcal{L}(\mathbf{t}_{b,m})$ under symmetries. We notice that all of them contain a BS factor as a vertex.

Step (ii) Proposition 4.2 says that every BS factor is of length $b^\ell n$, $\ell \in \mathbb{N}$, where $n \in \{1, \dots, 2b-1\}$. It is not difficult to see that all reduced Rauzy graphs of order greater than or equal to $2b$ containing a BS factor as their vertex are obtained by a repeated application of $\varphi_{b,m}$ simultaneously to vertices and edges of reduced Rauzy graphs of order n, where $2 \le n \le 2b-1$. By Proposition 4.2, the reduced Rauzy graph of order nb^ℓ obtained when $\varphi_{b,m}$ is applied ℓ times to the reduced Rauzy graph of order n, where $2 \le n \le 2b-1$, satisfies

$$\begin{aligned} \{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_{nb^\ell}\} &= \{\tfrac{1}{b^\ell}\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_n\}, \\ \{\rho(v) \mid v \text{ BS vertex in } \tilde{\Gamma}_{nb^\ell}\} &= \{\tfrac{1}{b^\ell}\rho(v) \mid v \text{ BS vertex in } \tilde{\Gamma}_n\}. \end{aligned}$$

Step (iii) Applying Observation 2.11, we obtain

(a) If $(n-1)b^\ell < N < nb^\ell$ for some $n \in \{2, \dots, 2b\}$, then

$$\begin{aligned} \{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{b,m})\} &= \{\tfrac{1}{b^\ell}\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_n\} \\ &\quad \cup \{\tfrac{1}{b^\ell}\rho(v) \mid v \text{ BS vertex in } \tilde{\Gamma}_n\}. \end{aligned}$$

(b) If $N = nb^\ell$ for some $n \in \{2, \dots, 2b-1\}$, then

$$\{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{b,m})\} = \{\tfrac{1}{b^\ell}\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_n\}.$$

Let us illustrate the RRG method for the Thue–Morse word $\mathbf{t}_{2,2}$.

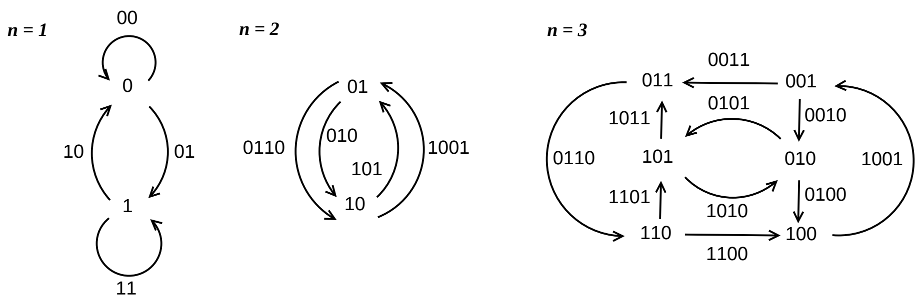

Fig. 1. Reduced Rauzy graphs of $\mathbf{t}_{2,2}$ of order $n \in \{1, 2, 3\}$.

Step (i) In the first step, we describe edge and vertex frequencies in $\tilde{\Gamma}_n$ for $n \in \{1, 2, 3\}$

$\tilde{\Gamma}_1$: $\rho(0) = \rho(1) = \frac{1}{2}$ and $\{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_1\} = \{\frac{1}{3}, \frac{1}{6}\}$.
Explanation:

- Thanks to Observation 3.2, we have $\rho(0) = \rho(1)$, $\rho(01) = \rho(10)$, and $\rho(00) = \rho(11)$.
- Using Property $a)$ of $\varphi_{b,m}$ and $\mathbf{t}_{b,m}$, we get $\rho(0) = \frac{1}{2}$.

- By Corollary 2.6 and Proposition 2.2, it holds $\rho(00) = \rho(1001) = \frac{1}{2}\rho(10)$.
- Applying the Kirchhoff's law for frequencies, we get $\rho(0) = \rho(01) + \rho(00) = \frac{3}{2}\rho(01)$, consequently $\rho(01) = \frac{1}{3}$.

$\tilde{\Gamma}_2$: $\rho(01) = \rho(10) = \frac{1}{3}$ and $\{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_2\} = \{\frac{1}{6}\}$.
Explanation:

- Thanks to Observation 3.2, we have $\rho(010) = \rho(101)$ and $\rho(0110) = \rho(1001)$.
- By Proposition 2.2, it holds $\rho(0110) = \frac{1}{2}\rho(01) = \frac{1}{6}$.
- Applying the Kirchhoff's law for frequencies, we get $\rho(01) = \rho(010) + \rho(0110)$. Therefore $\rho(010) = \frac{1}{6}$.

$\tilde{\Gamma}_3$: $\rho(010) = \rho(101) = \frac{1}{6}$ and $\{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_3\} = \{\frac{1}{6}, \frac{1}{12}\}$.
Explanation:

- Thanks to Observation 3.2, we have $\rho(011) = \rho(100) = \rho(001) = \rho(110)$, $\rho(0011) = \rho(1100)$, $\rho(0101) = \rho(1010)$, $\rho(0010) = \rho(1101) = \rho(1011) = \rho(0100)$.
- By Corollary 2.6 and Proposition 2.2, it holds $\rho(0010) = \rho(100101) = \frac{1}{2}\rho(100) = \frac{1}{2}\rho(1001) = \frac{1}{12}$ and $\rho(0011) = \rho(100110) = \frac{1}{2}\rho(101) = \frac{1}{12}$.
- The Kirchhoff's law for frequencies implies $\rho(0101) = \rho(010) - \rho(0100) = \frac{1}{12}$.

Step (ii) All reduced Rauzy graphs of order greater than or equal to 4 containing a BS factor as their vertex are depicted in Figure 2.

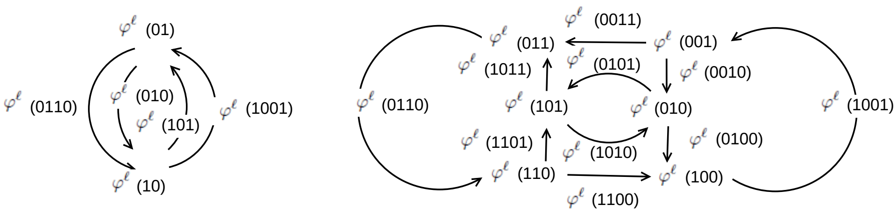

Fig. 2. For any reduced Rauzy graph of $\mathbf{t}_{2,2}$ of order $n \geq 4$ containing a BS factor as its vertex, there exists $\ell \geq 1$ such that the graph takes one of the depicted forms.

It holds for all $\ell \in \mathbb{N}$

$$\begin{aligned} \{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_{2\cdot 2^\ell}\} &= \{\tfrac{1}{2^\ell}\tfrac{1}{6}\}, \\ \{\rho(v) \mid v \text{ BS vertex in } \tilde{\Gamma}_{2\cdot 2^\ell}\} &= \{\tfrac{1}{2^\ell}\tfrac{1}{3}\}, \end{aligned}$$

and

$$\begin{aligned} \{\rho(e) \mid e \text{ edge in } \tilde{\Gamma}_{3\cdot 2^\ell}\} &= \{\tfrac{1}{2^\ell}\tfrac{1}{6}, \tfrac{1}{2^\ell}\tfrac{1}{12}\}, \\ \{\rho(v) \mid v \text{ BS vertex in } \tilde{\Gamma}_{3\cdot 2^\ell}\} &= \{\tfrac{1}{2^\ell}\tfrac{1}{6}\}. \end{aligned}$$

Step (iii) Applying the third step of the RRG method, we get the sets of factor frequencies $\{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{2,2})\}$ for $N \in \mathbb{N}$.

- $\{\rho(e) \mid e \in \mathcal{L}_1(\mathbf{t}_{2,2})\} = \{\frac{1}{2}\}$.
- $\{\rho(e) \mid e \in \mathcal{L}_2(\mathbf{t}_{2,2})\} = \{\frac{1}{3}, \frac{1}{6}\}$.
- If $2 \cdot 2^\ell < N < 3 \cdot 2^\ell$ for some $\ell \in \mathbb{N}$, then
$$\{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{2,2})\} = \left\{\frac{1}{2^\ell}\frac{1}{6}, \frac{1}{2^\ell}\frac{1}{12}\right\} \cup \left\{\frac{1}{2^\ell}\frac{1}{6}\right\} = \left\{\frac{1}{2^\ell}\frac{1}{6}, \frac{1}{2^\ell}\frac{1}{12}\right\}.$$
- If $3 \cdot 2^\ell < N < 4 \cdot 2^\ell$ for some $\ell \in \mathbb{N}$, then
$$\{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{2,2})\} = \left\{\frac{1}{2^{\ell+1}}\frac{1}{6}\right\} \cup \left\{\frac{1}{2^{\ell+1}}\frac{1}{3}\right\} = \left\{\frac{1}{2^{\ell+1}}\frac{1}{3}, \frac{1}{2^{\ell+1}}\frac{1}{6}\right\}.$$
- If $N = 2 \cdot 2^\ell$ for some $\ell \in \mathbb{N}$, then
$$\{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{b,m})\} = \left\{\frac{1}{2^\ell}\frac{1}{6}\right\}.$$
- If $N = 3 \cdot 2^\ell$ for some $\ell \in \mathbb{N}$, then
$$\{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{b,m})\} = \left\{\frac{1}{2^\ell}\frac{1}{6}, \frac{1}{2^\ell}\frac{1}{12}\right\}.$$

Application of the RRG method. The RRG method for $\mathbf{t}_{b,m}$ says that it suffices to describe frequencies of edges and vertices being BS factors in reduced Rauzy graphs of order n, where $1 \leq n \leq 2b-1$, in order to get $\{\rho(e) \mid e \in \mathcal{L}_{n+1}(e)\}$ for all $n \in \mathbb{N}$. Using the description of BS factors from [6], one can see that Step (i) of the RRG method has to be executed for $1 \leq n \leq b$ and for $b+1 \leq n \leq 2b-1$ separately.

Part 1)

For $1 \leq n \leq b$, the reduced Rauzy graph $\tilde{\Gamma}_n$ has m vertices. All of them are BS factors of the form $k(k+1)\dots(k+n-1)$. Since each of them is equal to $\Pi_k(01\dots(n-1))$, their frequencies are the same. Moreover,

- e is an edge ending in $01\dots(n-1)$ if and only if $\Pi_k(e)$ is an edge ending in $k(k+1)\dots(k+n-1)$,
- e is an edge ending in $01\dots(n-1)$ if and only if $\Psi_{k+n-1}(e)$ is an edge starting in $k(k+1)\dots\big(k+n-1)$,

and since $\rho(e) = \rho(\Pi_k(e)) = \rho(\Psi_{k+n-1}(e))$, it suffices to describe frequencies of edges ending in $01\dots(n-1)$ in order to get all edge frequencies of $\tilde{\Gamma}_n$. As shown in [6], $\mathrm{Lext}(01\dots(n-1)) = \{-1+k(b-1) \mid k \in \{0,1,\dots,q-1\}\}$, where q is taken from Definition 4.1.

Lemma 4.5. Consider the generalized Thue–Morse word $\mathbf{t}_{b,m}$ with $b \geq 2, m \geq 1, b, m \in \mathbb{N}$, and $b \not\equiv 1 \mod m$. Let q be the number given in Definition 4.1. Denote $f = \rho(01)$. Then $f = \frac{b^{q-1}}{m}\frac{b-1}{b^q-1}$ and for $1 \leq n \leq b$, the frequencies of the vertex $w = 01\ldots(n-1)$ and of the edges ending in w satisfy

$$\begin{array}{rcll} \rho\big(01\ldots(n-1)\big) &=& (n-1)f - \frac{n-2}{m} & \text{for } n \geq 1, \\ \rho\big((-1)01\ldots(n-1)\big) &=& nf - \frac{n-1}{m}, & \\ \rho\big((-1+k(b-1))01\ldots(n-1)\big) &=& \frac{1}{b^k}f & \text{for } k \in \{1,\ldots,q-1\}. \end{array}$$

P r o o f. It holds by Corollary 2.6 and Proposition 2.2 for $k \in \{1,\ldots,q-1\}$ that

$$\rho\big((-1+k(b-1))0\big) = \rho\big(\varphi(-1+(k-1)(b-1))0)\big) = \frac{1}{b}\rho\big((-1+(k-1)(b-1))0\big).$$

Thus $\rho\big((-1+k(b-1))0\big) = \frac{1}{b^k}\rho\big((-1)0\big) = \frac{1}{b^k}\rho\big(\Pi_{-1}(01)\big) = \frac{1}{b^k}f$. Using Observation 2.5, we obtain $f = \rho(0) - \sum_{k=1}^{q-1}\frac{f}{b^k}$. Therefore

$$f = \frac{b^{q-1}}{m}\frac{b-1}{b^q-1}.$$

Let us proceed by induction on n. Let $n = 1$, then $\rho(0) = \frac{1}{m}$ by Property $a)$ of $\varphi_{b,m}$ and $\mathbf{t}_{b,m}$. Let $1 < n+1 \leq b$. Assume

$$\begin{array}{rcll} \rho\big(01\ldots(n-1)\big) &=& (n-1)f - \frac{n-2}{m} & \text{for } n \geq 2, \\ \rho\big((-1)01\ldots(n-1)\big) &=& nf - \frac{n-1}{m}, & \\ \rho\big((-1+k(b-1))01\ldots(n-1)\big) &=& \frac{1}{b^k}f & \text{for } k \in \{1,\ldots,q-1\}. \end{array}$$

Then, $\rho(01\ldots n) = \rho\big(\Pi_{-1}(01\ldots n)\big) = \rho\big((-1)01\ldots(n-1)\big) = nf - \frac{n-1}{m}$. Applying Corollary 2.6, we get

$$\rho\big((-1+k(b-1))01\ldots n\big) = \rho\big((-1+k(b-1))01\ldots(n-1)\big) = \tfrac{1}{b^k}f.$$

Using the Kirchhoff's law for frequencies (Observation 2.5), we have

$$\rho\big((-1)01\ldots n\big) = nf - \tfrac{n-1}{m} - \textstyle\sum_{k=1}^{q-1}\tfrac{f}{b^k} = (n+1)f - \tfrac{n}{m}.$$ □

Part 2)

For $b+1 \leq n \leq 2b-1$, the reduced Rauzy graph $\tilde{\Gamma}_n$ has $3m$ vertices: m of them are BS factors of the form $\Pi_k(01\ldots(n-1))$, $k \in \mathbb{Z}_m$, m of them are LS factors (that are not RS factors) of the form $\Pi_k(01\ldots(b-1)1\ldots(n-b))$, $k \in \mathbb{Z}_m$, m of them are RS factors (that are not LS factors) obtained by applying Ψ_0 to LS factors. Since symmetries preserve frequencies, all BS factors have their frequency equal to $\rho(01\ldots(n-1))$ and similarly, all LS and RS factors have their frequency equal to $\rho(01\ldots(b-1)1\ldots(n-b))$. By analogous arguments as in part 1), we deduce that it suffices to describe frequencies of edges ending in $01\ldots(n-1)$ and in $01\ldots(b-1)1\ldots(n-b)$ and the frequency of the unique edge $01\ldots(b-1)1\ldots(n+1-b)$ starting in $01\ldots(b-1)1\ldots(n-b)$ in order to get all edge frequencies of $\tilde{\Gamma}_n$. Again by [6],

$$\begin{array}{rcl} \mathrm{Lext}(01\ldots(n-1)) &=& \{-1, b-2\}, \\ \mathrm{Lext}(01\ldots(b-1)1\ldots(n-b)) &=& \{-1+k(b-1) \mid k \in \{0,1,\ldots,q-1\}\}. \end{array}$$

Lemma 4.6. Consider the generalized Thue–Morse word $\mathbf{t}_{b,m}$ with $b \geq 2, m \geq 1, b, m \in \mathbb{N}$, and $b \not\equiv 1 \mod m$. Let q be the number given in Definition 4.1. Denote $f = \rho(01)$. Then for $b+1 \leq n \leq 2b-1$, the frequencies

1. of the BS vertex $w = 01\dots(n-1)$ and of the edges ending in w satisfy

$$\begin{aligned} \rho(01\dots(n-1)) &= \tfrac{1}{b^{q-1}} f - \tfrac{(n-b-1)}{b^q} f, \\ \rho\big((-1)01\dots(n-1)\big) &= \tfrac{1}{b^{q-1}} f - \tfrac{(n-b)}{b^q} f, \\ \rho\big((b-2)01\dots(n-1)\big) &= \tfrac{1}{b^q} f, \end{aligned}$$

2. of the edge $01\dots(b-1)1\dots(n+1-b)$ starting in the LS vertex $v = 01\dots(b-1)1\dots(n-b)$ and of the edges ending in v satisfy

$$\begin{aligned} \rho\big(01\dots(b-1)1\dots(n+1-b)\big) &= \tfrac{1}{b} f, \\ \rho\big((-1)01\dots(b-1)1\dots(n-b)\big) &= \tfrac{1}{b^q} f, \\ \rho\big((-1+(b-1))01\dots(b-1)1\dots(n-b)\big) &= \tfrac{1}{b}(2f - \tfrac{1}{m}), \\ \rho\big((-1+k(b-1))01\dots(b-1)1\dots(n-b)\big) &= \tfrac{1}{b^k} f \quad \text{for } k \in \{2,\dots,q-1\}. \end{aligned}$$

Proof. Let us proceed by induction on n.

(a) Let $n = b+1$, then using part 1), we obtain $\rho(01\dots b) = \rho\big(\Pi_{-1}(01\dots b)\big) = \rho\big((-1)01\dots(b-1)\big) = bf - \frac{b-1}{m} = \frac{1}{b^{q-1}} f$. By Corollary 2.6, Proposition 2.2, and Observation 3.2, we have $\rho\big((b-2)01\dots b\big) = \rho\big(\varphi((-1)0b)\big) = \rho\big(\varphi^2((1-b)b)\big) = \frac{1}{b^2}\rho\big((1-b)b\big) = \frac{1}{b^2}\rho\big(\Pi_{m-b}((1-b)b)\big) = \frac{1}{b^2}\rho\big((-1+(q-2)(b-1))0\big) = \frac{1}{b^q} f$. Finally, applying Observation 2.5, we get $\rho((-1)0\dots b) = \frac{1}{b^{q-1}} f - \frac{1}{b^q} f$.

Let $b+1 < n+1 \leq 2b-1$. Assume

$$\begin{aligned} \rho\big(01\dots(n-1)\big) &= \tfrac{1}{b^{q-1}} f - \tfrac{(n-b-1)}{b^q} f, \\ \rho\big((-1)01\dots(n-1)\big) &= \tfrac{1}{b^{q-1}} f - \tfrac{(n-b)}{b^q} f, \\ \rho\big((b-2)01\dots(n-1)\big) &= \tfrac{1}{b^q} f. \end{aligned}$$

Then, $\rho(01\dots n) = \rho\big(\Pi_{-1}(01\dots n)\big) = \rho\big((-1)01\dots(n-1)\big) = \frac{1}{b^{q-1}} f - \frac{(n-b)}{b^q} f$. Applying Corollary 2.6, we get $\rho\big((b-2)01\dots n\big) = \rho\big((b-2)01\dots(n-1)\big) = \frac{1}{b^q} f$. Using the Kirchhoff's law for frequencies (Observation 2.5), we have $\rho((-1)01\dots n) = \frac{1}{b^{q-1}} f - \frac{(n-b)}{b^q} f - \frac{1}{b^q} f = \frac{1}{b^{q-1}} f - \frac{(n+1-b)}{b^q} f$.

(b) Let $n = b+1$, then by Corollary 2.6 and Proposition 2.2, it follows $\rho(01\dots(b-1)12) = \rho(\varphi(01)) = \frac{1}{b} f$. Again, by Corollary 2.6 and Proposition 2.2, it holds for $k \in \{2,\dots,q-1\}$ that $\rho\big((-1+k(b-1))01\dots(b-1)1\big) = \rho\big(\varphi((-1+(k-1)(b-1))01)\big) = \frac{1}{b}\rho\big((-1+(k-1)(b-1))01\big) = \frac{1}{b^k} f$, and for $k = 1$, we have by the same arguments $\rho\big((-1+(b-1))01\dots(b-1)1\big) = \rho\big(\varphi((-1)01)\big) = \frac{1}{b}\rho\big((-1)01\big) = \frac{1}{b}(2f - \frac{1}{m})$. Finally, by the Kirchhoff's law for frequencies (Observation 2.5), we derive $\rho\big((-1)01\dots(b-1)1\big) = \frac{1}{b} f - \frac{1}{b}(2f - \frac{1}{m}) - \sum_{k=2}^{q-1} \frac{1}{b^k} f = \frac{1}{b^q} f$.

Let $b+1 < n+1 \leq 2b-1$. Assume

$$\begin{aligned} \rho\big(01\dots(b-1)1\dots(n+1-b)\big) &= \tfrac{1}{b} f, \\ \rho\big((-1)01\dots(b-1)1\dots(n-b)\big) &= \tfrac{1}{b^q} f, \\ \rho\big((-1+(b-1))01\dots(b-1)1\dots(n-b)\big) &= \tfrac{1}{b}(2f - \tfrac{1}{m}), \\ \rho\big((-1+k(b-1))01\dots(b-1)1\dots(n-b)\big) &= \tfrac{1}{b^k} f \quad \text{for } k \in \{2,\dots,q-1\}. \end{aligned}$$

By Corollary 2.6, we have $\rho\big(01\dots(b-1)1\dots(n+2-b)\big) = \rho\big(01\dots(b-1)1\dots(n+1-b)\big) = \frac{1}{b}f$. Again by Corollary 2.6, we get for all $k \in \{2,\dots,q-1\}$ that $\rho\big((-1+k(b-1)01\dots(b-1)1\dots(n+1-b)\big) = \rho\big((-1+k(b-1)01\dots(b-1)1\dots(n-b)\big) = \frac{1}{b^k}f$, and analogously, $\rho\big((-1+(b-1)01\dots(b-1)1\dots(n+1-b)\big) = \rho\big((-1+(b-1)01\dots(b-1)1\dots(n-b)\big) = \frac{1}{b}(2f-\frac{1}{m})$. Using the Kirchhoff's law for frequencies (Observation 2.5), we have $\rho\big((-1)01\dots(b-1)1\dots(n+1-b)\big) = \frac{1}{b}f - \frac{1}{b}(2f-\frac{1}{m}) - \sum_{k=2}^{q-1}\frac{1}{b^k}f = \frac{1}{b^q}f$. □

Theorem 4.7. Let $b \geq 2, m \geq 1, b, m \in \mathbb{N}$, and $b \not\equiv 1 \mod m$. Let $\mathbf{t}_{b,m}$ be the fixed point of the morphism $\varphi_{b,m}$ defined in (2). Let q be the number given in Definition 4.1 and let $f = \rho(01)$. Then the sets of factor frequencies take the following form for $N \in \mathbb{N}$.

N	$\{\rho(e) \mid e \in \mathcal{L}_{N+1}(\mathbf{t}_{b,m})\}$
0	$\frac{1}{m}$
1	$\frac{f}{b^k}$, where $k \in \{0,\dots,q-1\}$
$(n-1)b^\ell < N < nb^\ell,\ \ell \in \mathbb{N}$, where $n \in \{3,\dots,b\}$	$\frac{1}{b^\ell}\left((n-1)f - \frac{n-2}{m}\right), \frac{1}{b^\ell}\left(nf - \frac{n-1}{m}\right), \frac{1}{b^\ell}\left(\frac{1}{b^k}f\right)$, where $k \in \{1,\dots,q-1\}$
$(n-1)b^\ell < N < nb^\ell,\ \ell \in \mathbb{N}$, where $n \in \{b+1,\dots,2b-1\}$	$\frac{1}{b^\ell}\left(\frac{2b+1-n}{b^q}f\right), \frac{1}{b^\ell}\left(\frac{1}{b^{q-1}}f - \frac{(n-b)}{b^q}f\right), \frac{1}{b^{\ell+1}}(2f-\frac{1}{m}), \frac{1}{b^\ell}\left(\frac{1}{b^k}f\right)$, where $k \in \{1,\dots,q\}$
$(2b-1)b^\ell < N < 2b^{\ell+1},\ \ell \in \mathbb{N}$	$\frac{1}{b^{\ell+1}}\left(2f-\frac{1}{m}\right), \frac{1}{b^{\ell+1}}\left(\frac{1}{b^k}f\right)$, where $k \in \{0,\dots,q-1\}$
$nb^\ell,\ \ell \in \mathbb{N}$, where $n \in \{2,\dots,b\}$	$\frac{1}{b^\ell}\left(nf - \frac{n-1}{m}\right), \frac{1}{b^\ell}\left(\frac{f}{b^k}\right)$, where $k \in \{1,\dots,q-1\}$
$nb^\ell, \ell \in \mathbb{N}$, where $n \in \{b+1,\dots,2b-1\}$	$\frac{1}{b^\ell}\left(\frac{2b-n}{b^q}f\right), \frac{1}{b^{\ell+1}}\left(2f-\frac{1}{m}\right), \frac{1}{b^\ell}\left(\frac{1}{b^k}f\right)$, where $k \in \{1,\dots,q\}$

Proof. The statement is obtained when putting together Lemmas 4.5 and 4.6 and Steps (ii) and (iii) of the RRG method. □

5. UPPER BOUND ON FREQUENCIES

In the last section, let us show and explain that the optimal upper bound on the number of factor frequencies in infinite words whose language is invariant under more symmetries, here recalled as Theorem 3.3, is not reached for large enough n for any generalized Thue–Morse word $\mathbf{t}_{b,m}$ with $b \geq 2, m \geq 1, b, m \in \mathbb{N}$, and $b \not\equiv 1 \mod m$. Let q be the number given in Definition 4.1. First of all, the upper bound cannot be attained for $q > 2$: for any length n there exist LS factors w whose number of extensions equals $\#\mathrm{Rext}(w) = q$ (see Remark 4.3), consequently the estimate $\#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w\ LS\} \leq \mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})}\big(\#\mathrm{Lext}(w) - 1\big)$, used in the proof of Theorem 3.3, is too rough for $q > 2$.

Nevertheless, even in the case of $q = 2$, if we take $n = (2b-1)b^\ell$ for any $l \in \mathbb{N}$, then $\#\{\rho(e) \mid e \in \mathcal{L}_{\mathbf{t}_{b,m}}(n+1)\} \leq 3 = q+1$ by Theorem 4.7. By the description of factor complexity from [6], we have $\mathcal{C}(n+1) - \mathcal{C}(n) = qm = 2m$. It follows from Properties of $\varphi_{b,m}$ and $\mathbf{t}_{b,m}$ summarized in Section 4 that $\#G = 2m$ and the number of BS factors of length n is equal to m and is the same as the number of BS factors being θ-palindromes for some antimorphism $\Psi_x,\ x \in \mathbb{Z}_m$. Therefore, the upper bound from Theorem 3.3 is equal to $\frac{1}{2m}\Big(8m + 2m - m - m\Big) = 4$. Hence, for any $M \in \mathbb{N}$, the equality in the upper bound from Theorem 3.3 is not reached for all $n \geq M$.

Let us explain the reason. In the proof of Theorem 3.3, we have used the invariance of $\mathcal{L}(\mathbf{t}_{b,m})$ under symmetries in order to obtain the upper bound on the number of factor frequencies. However, some factors may have the same frequency for another reason. We observe as a direct consequence of Corollary 2.6 the following.

Observation 5.1. If w is a BS factor of an infinite word $\mathbf{u}$ such that for every $a \in \mathrm{Lext}(w)$, there exists a unique $b \in \mathrm{Rext}(w)$ satisfying $awb \in \mathcal{L}(\mathbf{u})$ (let us call such BS factors slim), then $\rho(aw) = \rho(awb) = \rho(wb)$.

For $n = 2b-1$, the BS factor of the form $w = 01\ldots(2b-2)$ is slim: w can be extended in only two ways, as $(b-2)w(2b-1)$ and as $(-1)wb$. Hence, $\rho((-1)w) = \rho(wb)$ even if these factors are not symmetric images of each other. Similarly, the BS factor $v = \varphi^{\ell}_{b,m}(w)$ of length $n = (2b-1)b^{\ell}$ is slim and $\rho(av) = \rho(vb)$, where a is the last letter of $\varphi^{\ell}_{b,m}(-1)$, i.e., $a = -1 + \ell(b-1)$. It holds again that av and vb are not symmetric images of one another.

ACKNOWLEDGEMENT

I would like to thank 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, and by the grant of the Grant Agency of the Czech Technical University in Prague grant No. SGS11/162/OHK4/3T/14.

(Received July 28, 2011)

REFERENCES

[1] J.-P. Allouche and J. Shallit: Sums of digits, overlaps, and palindromes. Discrete Math. Theoret. Comput. Sci. *4* (2000), 1–10.

[2] L. Balková: Factor frequencies in languages invariant under symmetries preserving factor frequencies. Integers - Electronic Journal of Combinatorial Number Theory 12 (2012), A36.

[3] M. Dekking: On the Thue–Morse measure. Acta Univ. Carolin. Math. Phys. *33* (1992), 35–40.

[4] A. Frid: On the frequency of factors in a D0L word. J. Automata, Languages and Combinatorics *3* (1998), 29–41.

[5] M. Queffélec: Substitution dynamical systems – Spectral analysis. Lecture Notes in Math. *1294* (1987).

[6] Š. Starosta: Generalized Thue–Morse words and palindromic richness. Kybernetika *48* (2012), 3, 361–370.

Ľubomíra Balková, Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2. Czech Republic.
e-mail: lubomira.balkova@fjfi.cvut.cz