Computing the k-binomial complexity of generalized Thue–Morse words

M. Golafshan[1], M. Rigo*[1], and M. A. Whiteland†[2]

[1]Department of Mathematics, University of Liège, Liège, Belgium,
{mgolafshan,m.rigo}@uliege.be
[2]Department of Computer Science, Loughborough University, Epinal Way, LE11 3TU Loughborough, Leicestershire, United Kingdom,
m.a.whiteland@lboro.ac.uk

Abstract

Two finite words are k-binomially equivalent if each subword (i.e., subsequence) of length at most k occurs the same number of times in both words. The k-binomial complexity of an infinite word is a function that maps the integer $n \geqslant 0$ to the number of k-binomial equivalence classes represented by its factors of length n.

The Thue–Morse (TM) word and its generalization to larger alphabets are ubiquitous in mathematics due to their rich combinatorial properties. This work addresses the k-binomial complexities of generalized TM words. Prior research by Lejeune, Leroy, and Rigo determined the k-binomial complexities of the 2-letter TM word. For larger alphabets, work by Lü, Chen, Wen, and Wu determined the 2-binomial complexity for m-letter TM words, for arbitrary m, but the exact behavior for $k \geqslant 3$ remained unresolved. They conjectured that the k-binomial complexity function of the m-letter TM word is eventually periodic with period m^k.

We resolve the conjecture positively by deriving explicit formulae for the k-binomial complexity functions for any generalized TM word. We do this by characterizing k-binomial equivalence among factors of generalized TM words. This comprehensive analysis not only solves the open conjecture, but also develops tools such as abelian Rauzy graphs.

Keywords: Thue–Morse word ; binomial coefficient of words ; binomial complexity ; Rauzy graph ; abelian equivalence.
2000 Mathematics Subject Classification: 68R15, 68Q45.

Contents

1 Introduction **2**
1.1 Previously known results on generalized Thue–Morse words 4
1.2 Main results 5

2 Key Points of Our Proof Strategy **7**

3 Compilation of Preliminary Results **9**

4 Ability to Discern k-Binomially Non-Equivalent Factors **10**

*Corresponding author. The first two authors are supported by the FNRS Research grant T.196.23 (PDR)
†Part of the work was performed while affiliated with Univeristy of Liège and supported by the FNRS Research grant 1.B.466.21F

5 Recognizability and Structure of Factors **13**

6 Counting Classes of a New Equivalence Relation **16**

7 Characterizing Binomial Equivalence in $\mathbf{t}_m$ **20**
7.1 Proof of Lemma 7.1 — The base case 21
7.2 Proof of Lemma 7.1 — The induction step 24

8 Abelian Complexity for Short Factors **28**

9 Description of the Abelian Rauzy Graphs **30**
9.1 When $\ell < m$ 31
9.2 When $\ell \geqslant m$ 34

10 Proof of Lemma 4.4 **37**

1 Introduction

The Thue–Morse infinite word (or sequence) $\mathbf{t}_2 = 011010011001\cdots$ is the fixed point of the morphism $\sigma_2 : 0 \mapsto 01, 1 \mapsto 10$ starting with 0. It was originally constructed by A. Thue in the context of avoidable patterns. It does not contain any overlap of the form $auaua$ where $a \in \{0,1\}$ and $u \in \{0,1\}^*$. This word was later rediscovered by M. Morse while studying differential geometry and geodesics on surfaces of negative curvature [20]. The study of non-repetitive structures is fundamental in combinatorics. See references [9, 15] for further details. The Thue–Morse word has found applications across a wide range of fields including mathematics, physics, economics, and computer science [2, 3]. In number theory, the word is linked to the Prouhet–Tarry–Escott problem [35]. Additionally, L. Mérai and A. Winterhof have analyzed its pseudo-random characteristics; see e.g., [19]. The Thue–Morse word also emerges in physics as an example of an aperiodic structure that exhibits a singular continuous contribution to the diffraction pattern [34, 14]. This property is significant in the study of quasi-crystals and materials with non-periodic atomic arrangements [30] or fractal geometry [13]. In economics or game theory, the Thue–Morse word has been proposed to ensure fairness in sequential tournament competitions between two agents [21].

Let $m \geqslant 2$ be an integer and $\mathcal{A}_m = \{0, \ldots, m-1\}$ be the alphabet identified with the additive group $\mathbb{Z}/(m\mathbb{Z})$. Hereafter, all operations on letters are considered modulo m. Hence notation such as $\pmod m$ will usually be omitted.

This article deals with a natural generalization of the Thue–Morse word over an alphabet of size $m \geqslant 2$. Our primary goal is to identify and count subwords occurring in this word. This process directly relates to the notion of binomial complexity. We consider the morphism $\sigma_m : \mathcal{A}_m^* \to \mathcal{A}_m^*$, defined by

$$\sigma_m : i \mapsto i(i+1)\cdots(i+m-1).$$

The images $\sigma_m(i)$ correspond to cyclic shifts of the word $012\cdots(m-1)$. For instance, σ_2 is the classical Thue–Morse morphism. Our focus is on the infinite words $\mathbf{t}_m := \lim_{j\to\infty} \sigma_m^j(0)$. For example, we have

$$\mathbf{t}_3 = 0121202011202010122010121 20\cdots.$$

Throughout this article, infinite words are denoted using boldface symbols. The Thue–Morse word $\mathbf{t}_2$ and its generalizations $\mathbf{t}_m$ play a prominent role in combinatorics on words [3]. It serves as an example of an m-automatic sequence, where each letter is mapped by the morphism σ_m to an image of uniform length m. Thus, σ_m is said to be m-*uniform*. Further results on subwords of $\mathbf{t}_m$ in arithmetic progressions can be found in [22].

In this article, we distinguish between a *factor* and a *subword* of a word $w = a_1 a_2 \cdots a_\ell$. A factor consists of consecutive symbols $a_i a_{i+1} \cdots a_{i+n-1}$, whereas a subword is a subsequence $a_{j_1} \cdots a_{j_n}$, with $1 \leqslant j_1 < \cdots < j_n \leqslant \ell$. Every factor is a subword, but the converse does not always hold. The set of factors of an infinite word $\mathbf{w}$ (respectively, factors of length n) is denoted by $\mathrm{Fac}(\mathbf{w})$ (respectively, $\mathrm{Fac}_n(\mathbf{w})$). Let $|x|$ (resp. $|x|_a$) denote the length of a finite word x (resp. the number of occurrences of a letter a in x). For general references on binomial coefficients of words and binomial equivalence, see [17, 23, 24, 25, 26].

Definition 1.1. Let u and w be words over a finite alphabet $\mathcal{A}$. Writing $u = a_1 \cdots a_n$, where $a_i \in \mathcal{A}$ for all i, the *binomial coefficient* $\binom{u}{w}$ is defined as

$$\binom{u}{w} = \#\left\{ i_1 < i_2 < \cdots < i_{|w|} \mid a_{i_1} a_{i_2} \cdots a_{i_{|w|}} = w \right\}.$$

It counts the number of occurrences of w as a subword of u.

Note that the same notation is used for the binomial coefficients of words and integers, as the context prevents any ambiguity (the binomial coefficient of unary words naturally coincides with the integer version: $\binom{a^n}{a^k} = \binom{n}{k}$).

Definition 1.2 ([25])**.** Two words $u, v \in \mathcal{A}^*$ are said to be k-*binomially equivalent*, and we write $u \sim_k v$, if

$$\binom{u}{x} = \binom{v}{x}, \quad \forall x \in \mathcal{A}^{\leqslant k}.$$

If u and v are not k-binomially equivalent, we write $u \not\sim_k v$.

A word u is a permutation of the letters in v if and only if $u \sim_1 v$. This relation is known as the *abelian equivalence*.

Definition 1.3. Let $k \geqslant 1$ be an integer. The k-*binomial complexity function* $\mathsf{b}_{\mathbf{w}}^{(k)} \colon \mathbb{N} \to \mathbb{N}$ for an infinite word $\mathbf{w}$ is defined as

$$\mathsf{b}_{\mathbf{w}}^{(k)} : n \mapsto \#(\mathrm{Fac}_n(\mathbf{w})/\sim_k).$$

For $k = 1$, the k-binomial complexity is nothing else but the *abelian complexity function*, denoted by $\mathsf{a}_{\mathbf{w}}(n)$. Note that, for all $n \geq 0$, $\mathsf{b}_{\mathbf{w}}^{(k)}(n) \leq \mathsf{b}_{\mathbf{w}}^{(k+1)}(n)$.

One reason that motivates the study of this complexity function is the following. M. Andrieu and L. Vivion have recently shown that the k-binomial complexity function is well-suited for studying hypercubic billiard words [6, 33]. These words encode the sequence of faces successively hit by a billiard ball in a d-dimensional unit cube. The ball moves in straight lines until it encounters a face, then bounces elastically according to the law of reflection. A notable property is that removing a symbol from a d-dimensional billiard word results in a $(d-1)$-dimensional billiard word. Consequently, the projected factors of the $(d-1)$-dimensional word are subwords of the d-dimensional word.

The connections between binomial complexity and Parikh-collinear morphisms are studied in [29].

Definition 1.4. Let $\Psi : \mathcal{B}^* \to \mathbb{N}^{\#\mathcal{B}}$, defined as $w \mapsto (|w|_{b_1}, \ldots, |w|_{b_m})$ be the Parikh map for a totally ordered alphabet $\mathcal{B} = \{b_1 < \cdots < b_m\}$. A morphism $\varphi \colon \mathcal{A}^* \to \mathcal{B}^*$ is said to be *Parikh-collinear* if, for all letters $a, b \in \mathcal{A}$, there exist constants $r_{a,b}, s_{a,b} \in \mathbb{N}$ such that $r_{a,b}\Psi(\varphi(b)) = s_{a,b}\Psi(\varphi(a))$. If $r_{a,b} = s_{a,b}$ for all $a, b \in \mathcal{A}$, the morphism is called *Parikh-constant*.

Proposition 1.5 ([29, Cor. 3.6])**.** *Let* $\mathbf{w}$ *denote a fixed point of a Parikh-collinear morphism. For any* $k \geqslant 1$*, there exists a constant* $C_k \in \mathbb{N}$ *satisfying* $\mathsf{b}_{\mathbf{w}}^{(k)}(n) \leqslant C_k$ *for all* $n \in \mathbb{N}$*.*

It is worth noting that the above proposition was previously stated for Parikh-constant fixed points in [25].

1.1 Previously known results on generalized Thue–Morse words

It is well-known that the factor complexity $p_{\mathbf{w}} : n \mapsto \#\operatorname{Fac}_n(\mathbf{w})$ of any automatic word $\mathbf{w}$, including the generalized Thue–Morse words, is in $\mathcal{O}(n)$. The factor complexity function of $\mathbf{t}_m$ is known exactly via results of Starosta [32]:

Theorem 1.6. *For any $m \geq 1$, we have $p_{\mathbf{t}_m}(0) = 1$, $p_{\mathbf{t}_m}(1) = m$, and*

$$p_{\mathbf{t}_m}(n) = \begin{cases} m^2(n-1) - m(n-2) & \text{if } 2 \leqslant n \leqslant m; \\ m^2(n-1) - m^{k+1} + m^k & \text{if } m^k + 1 \leqslant n \leqslant 2m^k - m^{k-1},\ k \geq 1; \\ m^2(n-1) - m^{k+1} + m^k + m\ell & \text{if } n = 2m^k - m^{k-1} + 1 + \ell, \\ & \text{with } 0 \leqslant \ell < m^{k+1} - 2m^k + m^{k-1},\ k \geq 1. \end{cases}$$

The abelian complexity of $\mathbf{t}_m$ is known to be ultimately periodic with period m, as established by Chen and Wen [8]. For example, $(a_{\mathbf{t}_2}(n))_{n\geqslant 0} = (1, 2, 3, 2, 3, \ldots)$ and $(a_{\mathbf{t}_3}(n))_{n\geqslant 0} = (1, 3, 6, 7, 6, 6, 7, 6, \ldots)$. Moreover, the period takes either two or three distinct values, depending on the parity of m, as described in the following result.

Theorem 1.7 ([8])**.** *Let $m \geqslant 2$ and $n \geqslant m$. Let $v = n \pmod m$.*

- *If m is odd, then we have*

$$a_{\mathbf{t}_m}(n) = \#(\operatorname{Fac}_n(\mathbf{t}_m)/\sim_1) = \begin{cases} \frac{1}{4}m(m^2-1) + 1, & \text{if } v = 0; \\ \frac{1}{4}m(m-1)^2 + m, & \text{otherwise.} \end{cases}$$

- *If m is even, then we have*

$$a_{\mathbf{t}_m}(n) = \begin{cases} \frac{1}{4}m^3 + 1, & \text{if } v = 0; \\ \frac{1}{4}m(m-1)^2 + \frac{5}{4}m, & \text{if } v \neq 0 \text{ is even}; \\ \frac{1}{4}m^2(m-2) + m, & \text{if } v \text{ is odd}. \end{cases}$$

It is important to note that the abelian complexity function of a word generated by a Parikh-collinear morphism is not always eventually periodic [27]. Furthermore, [28] shows that the abelian complexity function of such a word is automatic in the sense defined by Allouche and Shallit [5].

According to Proposition 1.5 the k-binomial complexity of $\mathbf{t}_m$ is bounded by a constant (that depends on k). Explicit expressions of the functions $b_{\mathbf{t}_2}^{(k)}$ have been established:

Theorem 1.8 ([16, Thm. 6])**.** *Let $k \geqslant 1$. For every length $n \geqslant 2^k$, the k-binomial complexity $b_{\mathbf{t}_2}^{(k)}(n)$ is given by*

$$b_{\mathbf{t}_2}^{(k)}(n) = 3 \cdot 2^k + \begin{cases} -3, & \text{if } n \equiv 0 \pmod{2^k}; \\ -4, & \text{otherwise.} \end{cases}$$

If $n < 2^k$, the k-binomial complexity $b_{\mathbf{t}_2}^{(k)}(n)$ is equal to the factor complexity $p_{\mathbf{t}_m}(n)$.

Let us also mention that infinite recurrent words where all factors appear infinitely often, sharing the same j-binomial complexity as the Thue–Morse word $\mathbf{t}_2$, for all $j \leqslant k$, have been characterized in [29].

The authors of [16] conclude that "... the expression of a formula describing the k-binomial complexity of $\mathbf{t}_m$ ($m > 2$) seems to be more intricate. Therefore, a sharp description of the constants related to a given Parikh-constant morphism appears to be challenging".

Indeed, the difficulty in obtaining such an expression already becomes apparent with the 2-binomial complexity. In [18], Lü, Chen, Wen, and Wu derived a closed formula for the 2-binomial complexity of $\mathbf{t}_m$.

Theorem 1.9 ([18, Thm. 2]). *For every length $n \geqslant m^2$ and alphabet size $m \geqslant 3$, the 2-binomial complexity $b^{(2)}_{t_m}(n)$ is given by*

$$b^{(2)}_{t_m}(n) = \begin{cases} a_{t_m}(n/m) + m(m-1)(m(m-1)+1), & \text{if } n \equiv 0 \pmod{m}; \\ m^4 - 2m^3 + 2m^2, & \text{otherwise}. \end{cases}$$

The authors of [18] propose the conjecture that, for all $k \geqslant 3$, the k-binomial complexity of the generalized Thue–Morse word $\mathbf{t}_m$ is ultimately periodic. Precisely,

Conjecture 1.10 ([18, Conj. 1]). *For every $k \geqslant 3$, the k-binomial complexity $b^{(k)}_{t_m}$ of the generalized Thue–Morse word is ultimately periodic with period m^k.*

In this article, we confirm this conjecture by getting the exact expression for the k-binomial complexity of $\mathbf{t}_m$ for alphabet of any size m.

A natural question, which was raised by the two referees of this article, is the following: why do we restrict ourselves to words $\mathbf{t}_m$ when a larger generalization was considered, for instance, in [7, 32]? The answer to this question is given in the following remark.

Remark 1.11. One can be tempted to extend further the family of Thue–Morse words by considering two integer parameters $m, b \geq 2$ and the j^{th} term of $\mathbf{t}_{b,m}$ is equal to the b-ary sum-of-digits of $j \geqslant 0$, reduced modulo m. In that case, $\mathbf{t}_{b,m}$ is a fixed point of the b-uniform morphism $\sigma_{b,m} : i \mapsto i(i+1)\cdots(i+b-1)$ over $\mathcal{A}_m^*$.

When $m = b$, we have $\mathbf{t}_{b,m} = \mathbf{t}_m$ which is a fixed point of a Parikh-constant morphism and, for all $k \geq 1$, its k-binomial complexity is thus bounded using Proposition 1.5. We have a similar situation for $b \equiv 0 \pmod{m}$, the morphism generating $\mathbf{t}_{b,m}$ is also Parikh-constant, e.g., $0 \mapsto 0101$ and $1 \mapsto 1010$ for $b = 4$ and $m = 2$.

When $b \equiv 1 \pmod{m}$, the word $\mathbf{t}_{b,m}$ is trivially periodic and the situation is uninteresting.

However, when m is not congruent to 0, 1 or -1 modulo m, the situation changes drastically. The matrix associated with $\sigma_{b,m}$ is circulant and has a dominating eigenvalue equal to b. It is an easy exercise to see that the second largest modulus for the eigenvalues is equal to $|\sin(b\pi/m)|/\sin(\pi/m)$ which is larger than 1. Adamczewski showed that in such a situation, the abelian complexity [1] and thus the k-binomial complexity is unbounded. This is the reason why we only consider the case $b = m$. One might expect to get similar results using similar techniques for $b \equiv 0 \pmod{m}$. If the abelian complexity is unbounded, none of the above results we aim to generalize may hold. For $b \equiv -1 \pmod{m}$, there is the issue that the second-most significant eigenvalue is 1, Adamczewski's techniques need to be investigated.

So the word $\mathbf{t}_m$ essentially captures the periodic behavior of the binomial complexity.

1.2 Main results

Let $k \geqslant 2$ and $m \geqslant 2$. The behavior of $b^{(k)}_{t_m}(n)$ depends on the length n of the factors and is fully characterized by the following three results. Recall Theorem 1.6 for an explicit expression for $p_{t_m}(n)$.

Theorem 1.12. *The length of the shortest factor which is k-binomially equivalent to a distinct factor is $2m^{k-1}$. In particular, for any length $n < 2m^{k-1}$, the k-binomial complexity $b^{(k)}_{t_m}(n)$ coincides with the factor complexity $p_{t_m}(n)$.*

In the following statement, we make use of the quantity

$$\#E_m(\nu) = \begin{cases} m(1 + \nu m - \nu), & \text{if } \nu < m; \\ m^3 - m^2 + m, & \text{otherwise} \end{cases}$$

which counts the number of edges of the abelian Rauzy graph of $\mathbf{t}_m$ of order ν. For the sake of presentation, this graph will be described later on in Definition 2.6.

Theorem 1.13. *Let* $n \in [2m^{k-1}, 2m^k)$.

1. *If* $n = \nu\, m^{k-1}$ *for some* $\nu \in \{2, \dots, 2m-1\}$, *then*

$$b^{(k)}_{\mathbf{t}_m}(\nu\, m^{k-1}) = (m^{k-1}-1)\#E_m(\nu) + a_{\mathbf{t}_m}(\nu).$$

2. *If* $n = \nu\, m^{k-1} + \mu$ *for some* $\nu \in \{2, \dots, 2m-1\}$ *and* $0 < \mu < m^{k-1}$, *then*

$$b^{(k)}_{\mathbf{t}_m}(\nu\, m^{k-1} + \mu) = (\mu-1)\#E_m(\nu+1) + (m^{k-1}-\mu-1)\#E_m(\nu) + \#Y_m(\nu)$$

where

$$\#Y_m(\nu) = \begin{cases} 2m(1+\nu m-\nu) - m\nu(\nu-1), & \text{if } \nu < m; \\ m^3 - m^2 + 2m, & \text{otherwise.} \end{cases}$$

Theorem 1.14. *For every length* $n \geqslant 2m^k$, *if* $\lambda = n \pmod{m^k}$ *and* $\lambda = \nu m^{k-1} + \mu$ *with* $\nu < m$ *and* $\mu < m^{k-1}$, *we have*

$$b^{(k)}_{\mathbf{t}_m}(n) = (m^{k-1}-1)(m^3 - m^2 + m) + \begin{cases} a_{\mathbf{t}_m}(m+\nu), & \text{if } \mu = 0; \\ m, & \text{otherwise.} \end{cases}$$

In particular, $\left(b^{(k)}_{\mathbf{t}_m}(n)\right)_{n \geqslant 2m^k}$ *is periodic with period* m^k.

Combining the above two theorems, we conclude that the periodic part of $b^{(k)}_{\mathbf{t}_m}(n)$ begins at m^k and therefore answer positively to Conjecture 1.10. Periodicity for $n \geq 2m^k$ is given by Theorem 1.14 and is easily extended for $n \geq m^k$ by rewriting the expressions given in Theorem 1.13 for $n \in [m^k, 2m^k)$.

Corollary 1.15. *The sequence* $\left(b^{(k)}_{\mathbf{t}_m}(n)\right)_{n \geqslant m^k}$ *is periodic with period* m^k.

Example 1.16. Fig. 1 illustrates the 2- and 3-binomial complexities of $\mathbf{t}_3$. For short lengths, as described by Theorem 1.12, the factor complexity is shown using a black dashed line, while values from Theorem 1.13 are depicted in yellow. For larger lengths, values given by Theorem 1.14 are shown in purple and blue, with one period over $[2m^k, 3m^k)$ highlighted in purple.

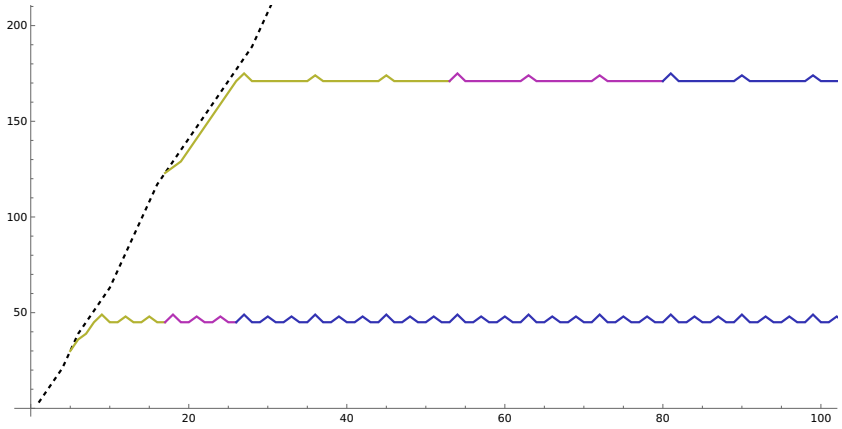


Figure 1: The first few values of the factor complexity (dashed), 2-binomial (below), and 3-binomial (above) complexities of $\mathbf{t}_3$.

For $m = 3$ and $k = 2, \dots, 6$, Table 1 provides words of length m^k realizing the period of the k-binomial complexity of $\mathbf{t}_3$, where exponents denote repetitions of an element.

k	
2	$49, 45^2, 48, 45^2, 48, 45^2$
3	$175, 171^8, 174, 171^8, 174, 171^8$
4	$553, 549^{26}, 552, 549^{26}, 552, 549^{26}$
5	$1687, 1683^{80}, 1686, 1683^{80}, 1686, 1683^{80}$
6	$5089, 5085^{242}, 5088, 5085^{242}, 5088, 5085^{242}$

Table 1: The period of $\mathsf{b}_{\mathbf{t}_3}^{(k)}$ for $k = 2, \ldots, 6$.

Let us highlight that Theorem 1.14 simultaneously generalizes the results from [16] and [18]. Furthermore, for $k = 2$, our formula reduces to Theorem 1.9. We also compute the values of $\mathsf{b}_{\mathbf{t}_m}^{(2)}(n)$ for the short lengths $n < m^2$. For $m = 2$, Theorem 1.14 provides the following result. For every length $n \geqslant 2^k$, we have:

$$\mathsf{b}_{\mathbf{t}_2}^{(k)}(n) = 3 \cdot 2^k + \begin{cases} -6 + \mathsf{a}_{\mathbf{t}_m}(2), & \text{if } n \equiv 0 \pmod{2^k}; \\ -6 + \mathsf{a}_{\mathbf{t}_m}(3), & \text{otherwise.} \end{cases}$$

This result corresponds to Theorem 1.8, with the shortest factors being handled by Theorem 1.12.

2 Key Points of Our Proof Strategy

The developments presented are relatively intricate. Therefore, we schematically outline the main steps of the proof. We hope this provides the reader with a general understanding of the structure of the article, allowing each section to be read almost independently of the others. This, we believe, makes the article easier to follow.

Definition 2.1. Let $j \geqslant 1$ and u be a factor of $\mathbf{t}_m$. A factorization of the form $u = x\sigma_m^j(u)y$ is referred to as a σ_m^j-*factorization* if there exists a factor aub of $\mathbf{t}_m$, where $a, b \in \mathcal{A}_m \cup \{\varepsilon\}$. In this factorization, x (respectively, y) must be a proper suffix (respectively, prefix) of $\sigma_m^j(a)$ (respectively, $\sigma_m^j(b)$). Here, ε is regarded as both a proper prefix and a proper suffix of itself.

In the literature, the terms *interpretation in* $\mathbf{t}_m$ and *ancestor* are also used. See, for instance, [11].

Theorem 1.14 addresses long enough factors. As discussed in Section 5, any factor $u \in \mathrm{Fac}(\mathbf{t}_m)$ of length $\geqslant 2m^k$ has a unique σ_m^k-factorization of the form $p_u\, \sigma_m^k(u) s_u$ (see Corollary 5.6). In particular, notice that $|p_u|, |s_u| < m^k$. Thus, we can associate each such factor u with a unique pair (p_u, s_u), leading to the following definition. As explained below, instead of studying the k-binomial equivalence $\sim_k$, we can instead consider another equivalence relation.

Definition 2.2. The equivalence relation on $\mathcal{A}_m^{<m^k} \times \mathcal{A}_m^{<m^k}$ is defined by $(p_1, s_1) \equiv_k (p_2, s_2)$ if there exist $x, y, p, q, r, t \in \mathcal{A}_m^*$ satisfying $|x|, |y| < m^{k-1}$ and

$$\begin{aligned} (p_1, s_1) &= \big(x\sigma_m^{k-1}(p), \sigma_m^{k-1}(q)y\big), \\ (p_2, s_2) &= \big(x\sigma_m^{k-1}(r), \sigma_m^{k-1}(t)y\big), \end{aligned}$$

and one of the following conditions holds

- $pq \sim_1 rt$,
- $pq \sim_1 rt\, \sigma_m(0)$,
- $pq\, \sigma_m(0) \sim_1 rt$.

We will show the following result in Section 4.

Proposition 2.3. *Let* $x, y \in \mathcal{A}_m^*$ *and* $k \geqslant 1$. *Then,* $x \sim_1 y$ *holds if and only if* $\sigma_m^k(x) \sim_{k+1} \sigma_m^k(y)$.

To achieve this result, a key challenge was identifying a suitable subword z of length $k+1$ such that $x \not\sim_1 y$, implies $\binom{x}{z} \neq \binom{y}{z}$. Section 4 focuses on providing the necessary computations to distinguish non-equivalent factors.

It can easily be shown that if $U, V \in \mathrm{Fac}(\mathbf{t}_m)$ are factors of length at least $2m^k$ and $(p_U, s_U) \equiv_k (p_V, s_V)$, then $U \sim_k V$. See Proposition 6.1. Moreover, the converse of this property is also valid. However, further developments, as outlined below, are necessary to prove this result.

Assuming, for now, that $(p_U, s_U) \equiv_k (p_V, s_V)$ if and only if $U \sim_k V$, proving Theorem 1.14, reduces to counting the number

$$\#\{(p_U, s_V) \mid U \in \mathrm{Fac}_n(\mathbf{t}_m)\}/\equiv_k$$

of such equivalence classes for $n \geqslant 2m^k$. This forms the core of Section 6 and is given by Theorem 6.5, whose statement is similar to Theorem 1.14.

To prove that $U \sim_k V$ implies $(p_U, s_U) \equiv_k (p_V, s_V)$, we first obtain the generalization of [18, Thm. 2] originally stated for 2-binomial equivalence. This result is then extended to all $k \geqslant 2$.

Proposition 2.4. *Let $k \geqslant 2$. For any two factors U and V of $\mathbf{t}_m$, the relation $U \sim_k V$ holds if and only if there exist σ_m^{k-1}-factorizations $U = p_U \sigma_m^{k-1}(u) s_U$ and $V = p_V \sigma_m^{k-1}(v) s_V$, such that $p_U = p_V$, $s_U = s_V$, and $u \sim_1 v$.*

We proceed by induction on k. The base case for $k = 2$ is essentially [18, Thm. 2]. However, our result slightly improves upon that of Chen et al. by not requiring any assumptions about the lengths of U and V in the factorizations.

Using Proposition 2.4, we can easily deduce the following result, thereby concluding this part.

Proposition 2.5. *Let $k \geqslant 2$. Let U and V be factors of $\mathbf{t}_m$ with the same length $\geqslant 2m^k$ such that*

$$U = p_U \sigma_m^{k-1}(\alpha_u\, \sigma_m(u) \beta_u) s_U, \quad \text{and} \quad V = p_V \sigma_m^{k-1}(\alpha_v\, \sigma_m(v) \beta_v) s_V,$$

where $|p_U|, |s_U|, |p_V|, |s_V| < m^{k-1}$ and $|\alpha_u|, |\beta_u|, |\alpha_v|, |\beta_v| < m$. If $U \sim_k V$, then

$$\big(p_U \sigma_m^{k-1}(\alpha_u), \sigma_m^{k-1}(\beta_u) s_U\big) \equiv_k \big(p_V \sigma_m^{k-1}(\alpha_v), \sigma_m^{k-1}(\beta_v) s_V\big).$$

We now focus on factors of length $n \in [2m^{k-1}, 2m^k)$. The proof of Theorem 1.13 relies on analyzing the so-called abelian Rauzy graphs.

Definition 2.6. For an infinite word, the *abelian Rauzy graph* of order $\ell \geqslant 1$ is defined with vertices corresponding to the abelian equivalence classes of factors of length ℓ (or equivalently, to their Parikh vectors). The edges of the graph are defined as follows. Let a, b be letters. If aUb is a factor of length $\ell + 1$, there exists a directed edge from $\Psi(aU)$ to $\Psi(Ub)$ labeled (a, b).

Let $G_{m,\ell}$ denote the abelian Rauzy graph of order ℓ of $\mathbf{t}_m$. The number of vertices in $G_{m,\ell}$ is clearly $\mathsf{a}_{\mathbf{t}_m}(\ell)$. For all $\ell \geqslant 1$, we define the following sets:

$$\begin{aligned} Y_{m,R}(\ell) &:= \{(\Psi(U), a) \mid a \in \mathcal{A}_m,\, Ua \in \mathrm{Fac}_{\ell+1}(\mathbf{t}_m)\}, \\ Y_{m,L}(\ell) &:= \{(a, \Psi(U)) \mid a \in \mathcal{A}_m,\, aU \in \mathrm{Fac}_{\ell+1}(\mathbf{t}_m)\}, \\ Y_m(\ell) &:= Y_{m,R}(\ell) \cup Y_{m,L}(\ell). \end{aligned}$$

Since $\mathbf{t}_m = \sigma_m^{k-1}(\mathbf{t}_m)$, it is quite straightforward to adapt [29, Prop. 5.5]. The idea behind the following formula is that to get $\mathsf{b}_{\mathbf{t}_m}^{(k)}(j\, m^{k-1} + r)$, one has to count the distinct σ_m^{k-1}-factorizations up to the equivalence relation given by Proposition 2.4.

Proposition 2.7. *Let $k \geqslant 2$. We let $E_m(j)$ denote the set of edges in the abelian Rauzy graph $G_{m,j}$. For all $j \geqslant 2$ and $0 < r < m^{k-1}$, the following holds*

$$\mathsf{b}_{\mathbf{t}_m}^{(k)}\big(j\, m^{k-1}\big) = \big(m^{k-1} - 1\big)\, \#E_m(j) + \mathsf{a}_{\mathbf{t}_m}(j),$$

and

$$\mathsf{b}_{\mathbf{t}_m}^{(k)}\big(j\, m^{k-1} + r\big) = (r-1)\, \#E_m(j+1) + (m^{k-1} - r - 1)\, \#E_m(j) + \#Y_m(j).$$

The reader may notice that the formula leading to Theorem 1.13 requires the values of the abelian complexity for short factors. However, Theorem 1.7 provides these values only for $j \geqslant m$, leaving the case $j < m$ unaddressed. Therefore, in Section 8, we describe the missing values of $a_{\mathbf{t}_m}(j)$ for $j < m$. In Section 9, we proceed to a detailed analysis of the structure of the abelian Rauzy graph of order j. We are thus able to determine explicit expressions for $\#E_m(j)$ and $\#Y_m(j)$.

3 Compilation of Preliminary Results

For the sake of completeness, we recall some basic properties of binomial coefficients [17, 25], which are implicitly applied throughout this article. Recall that $\binom{x}{\varepsilon} = 1$.

Lemma 3.1. *Let x, y, z be three words over the alphabet $\mathcal{A}$. The following relation holds*

$$\binom{xy}{z} = \sum_{\substack{u,v\in\mathcal{A}^* \\ uv=z}} \binom{x}{u}\binom{y}{v}.$$

More generally, let $x_1, \ldots, x_\ell$, $z \in \mathcal{A}^*$ and $\ell \geqslant 1$. Then, the following relation holds

$$\binom{x_1 \cdots x_\ell}{z} = \sum_{\substack{e_1,\ldots,e_\ell\in\mathcal{A}^* \\ e_1\cdots e_\ell=z}} \prod_{i=1}^{\ell} \binom{x_i}{e_i}.$$

Lemma 3.2 (Cancellation property)**.** *Let u, v, w be three words. The following equivalences hold*

- *$v \sim_k w$ if and only if $uv \sim_k uw$;*
- *$v \sim_k w$ if and only if $vu \sim_k wu$.*

We present a few straightforward observations regarding generalized Thue–Morse words. See, for instance, [31].

Proposition 3.3 ([4, Thm. 1])**.** *For any $m \geqslant 2$, the word $\mathbf{t}_m$ is overlap-free.*

The next statement is an easy observation.

Lemma 3.4. *Let $i, j \in \mathcal{A}_m$. If $i < j$ (respectively, $i > j$), the word ij appears exactly once as a subword in $m - j + i$ (respectively, $i - j$) of the images $\sigma_m(0), \sigma_m(1), \ldots, \sigma_m(m-1)$. Furthermore, the word ii does not occur as a subword in any of these images. Conversely, the $\binom{m}{2}$ distinct 2-subwords appearing in $\sigma_m(j)$ are given by $(j+t)(j+t+r)$, for $t = 0, \ldots, m-2$ and $r = 1, \ldots, m-t-1$.*

Let $\tau_m \colon \mathcal{A}_m^* \to \mathcal{A}_m^*$ be the cyclic morphism where each letter $a \in \mathcal{A}_m$ is mapped to $a + 1$. Because the compositions $\sigma_m \circ \tau_m$ and $\tau_m \circ \sigma_m$ are equal, the following lemma holds.

Lemma 3.5 ([32, Prop. 3.1])**.** *For all $n \geqslant 1$, the set $\mathrm{Fac}_n(\mathbf{t}_m)$ is closed under τ_m.*

The following result, proven in [8, Lem. 2], uses the concept of boundary sequence introduced in [12].

Lemma 3.6. *For all letters $a, b \in \mathcal{A}_m$ and all integer $n \geqslant 0$, there exists a factor of $\mathbf{t}_m$ in the form awb, where $|w| = n$. In particular, $\mathrm{Fac}_2(\mathbf{t}_m) = \mathcal{A}_m^2$.*

Since σ_m is Parikh-constant, the following result holds.

Proposition 3.7. *Assume $k \geqslant 1$. For all $u, v \in \mathcal{A}_m^*$, the following hold.*

(i) If $u \sim_k v$, then $\sigma_m(u) \sim_{k+1} \sigma_m(v)$.

(ii) If $u \sim_1 v$, then $\sigma_m^k(u) \sim_{k+1} \sigma_m^k(v)$.

(iii) If $|u| = |v|$, then $\sigma_m^k(u) \sim_k \sigma_m^k(v)$.

Proof. The first two statements are direct consequences of [29, Prop. 3.9], which applies to any Parikh-collinear morphism. For all letters $i, j \in \mathcal{A}_m$, it holds that $\sigma_m(i) \sim_1 \sigma_m(j)$. Hence, if two words u and v have the same length, then $\sigma_m(u) \sim_1 \sigma_m(v)$. So statement (iii) follows directly from statement (ii). □

4 Ability to Discern k-Binomially Non-Equivalent Factors

The purpose of this section is to express differences of the form $\binom{\sigma_m^k(u)}{x} - \binom{\sigma_m^k(v)}{x}$ for suitable subwords x. We additionally compute $\binom{\sigma_m^k(u)}{x} - \binom{\sigma_m^k(u)}{y}$ for an appropriate choice of x and y.

Recall the convention that $\mathcal{A}_m = \mathbb{Z}/(m\mathbb{Z})$, meaning any $i \in \mathbb{Z}$ is replaced with $(i \bmod m)$. For example, a letter (-1) is identified with $m-1$. For convenience, if $a \in \mathbb{N}$, we let $\overline{a}$ denote $-a$. As an example, with $m = 4$, the expression $2(-3)4(-1) = 2\overline{3}0\overline{1}$ is indeed 2103. In particular, the word $0\overline{1}\cdots\overline{k}$, which has length $k+1$, is a prefix of the periodic word $(0\overline{1}\,\overline{2}\cdots 1)^\omega$.

This kind of result is particularly useful for proving that two factors are not $(k+1)$-binomially equivalent.

Proposition 4.1. *Let $m \geqslant 2$, $k \geqslant 1$ and $i \in \mathcal{A}_m$. Then for all $j \in \mathcal{A}_m \setminus \{i\}$, the following holds*

$$\binom{\sigma_m^k(i)}{i(i-1)\cdots(i-k)} - \binom{\sigma_m^k(j)}{i(i-1)\cdots(i-k)} = m^{\binom{k}{2}}.$$

In particular, the coefficients $\binom{\sigma_m^k(j)}{i(i-1)\cdots(i-k)}$ are identical for all $j \neq i$.

For the classical Thue–Morse morphism, note that $\overline{1} = 1$ since $m = 2$. As an example, we have:

$$\binom{\sigma_2^{2n}(0)}{(01)^n 0} - \binom{\sigma_2^{2n}(1)}{(01)^n 0} = 2^{n(2n-1)}$$

and

$$\binom{\sigma_2^{2n+1}(0)}{(01)^{n+1}} - \binom{\sigma_2^{2n+1}(1)}{(01)^{n+1}} = 2^{n(2n+1)}.$$

Proof. By Lemma 3.5, we may assume $i = 0$ and it is enough to prove that

$$\binom{\sigma_m^k(0)}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^k(j)}{0\overline{1}\cdots\overline{k}} = m^{\binom{k}{2}}.$$

We proceed by induction on k. For the base case $k = 1$, Lemma 3.4 shows that the subword $0\overline{1}$ occurs exactly once in $\sigma_m(0)$ and does not appear in any other $\sigma_m(j)$ for $j \neq 0$. Assume that the statement holds for some $k \geqslant 1$. We now prove it for $k+1$.

The word $u = \sigma_m^{k+1}(0)$ can be factorized into m consecutive words, each of length m^k (referred to as m^k-*blocks*), as follows: $u = \sigma_m^k(0)\sigma_m^k(1)\cdots\sigma_m^k(\overline{1})$. The word $v = \sigma_m^{k+1}(j)$ is a cyclic permutation of the m^k-blocks of u, given by

$$v = \sigma_m^k(j)\cdots\sigma_m^k(\overline{1})\sigma_m^k(0)\cdots\sigma_m^k(j-1).$$

Our task is to count (or at least compare, as we are only interested in the difference) the occurrences of subwords $w = 0\overline{1}\cdots\overline{k}\,\overline{k+1}$ of length $k+2$ in u and v.

First, the number of occurrences fully contained within a single m^k-block is identical in u and v because they have the same m^k-blocks.

Next, we count the occurrences of w that are split across more than one m^k-block. These occurrences can be categorized into two cases:

I) w is split across at least two blocks, with no more than k letters of w appearing in each m^k-block. Proposition 3.7 ensures that $\sigma_m^k(i) \sim_k \sigma_m^k(i')$ for all letters i and i'. So u and v contain the same number of these types of occurrences.

II) w is split across at least two blocks, with $k+1$ letters of w appearing within a single m^k-block.

A difference arises only when $k+1$ letters of w appear within a single m^k-block, while its first or last letter belongs to a different m^k-block. By induction hypothesis, $\binom{\sigma_m^k(i)}{0\overline{1}\cdots\overline{k}} = \binom{\sigma_m^k(i')}{0\overline{1}\cdots\overline{k}}$ for any $i, i' \neq 0$. Similarly, $\binom{\sigma_m^k(i)}{\overline{1}\cdots\overline{k+1}} = \binom{\sigma_m^k(i')}{\overline{1}\cdots\overline{k+1}}$ for $i, i' \neq \overline{1}$. So to get different contributions, we only focus on the occurrences of the blocks $\sigma_m^k(0)$ and $\sigma_m^k(\overline{1})$ in u and v.

Let us first consider $\sigma_m^k(0)$. It appears at the beginning of u and it contains the subword $0\overline{1}\cdots\overline{k}$ exactly $\binom{\sigma_m^k(0)}{0\overline{1}\cdots\overline{k}}$ times. Moreover $\overline{k+1}$ occurs once in every of the subsequent $(m-1)m^{k-1}$ blocks of length m within $\sigma_m^k(1)\cdots\sigma_m^k(\overline{1})$. However, the first m^k-block in v is $\sigma_m^k(j)$, where the subword $0\overline{1}\cdots\overline{k}$ appears only $\binom{\sigma_m^k(j)}{0\overline{1}\cdots\overline{k}}$ times. By induction hypothesis, the resulting difference is

$$m^{\binom{k}{2}}(m-1)m^{k-1}.$$

A similar reasoning applies to $\sigma_m^k(\overline{1})$, which appears as the suffix of u and contains the subword $\overline{1}\,\overline{2}\cdots\overline{k+1}$ exactly $\binom{\sigma_m^k(\overline{1})}{\overline{1}\,\overline{2}\cdots\overline{k+1}}$ times. Moreover, 0 occurs exactly once in each of the preceding $(m-1)m^{k-1}$ blocks of length m within $\sigma_m^k(0)\cdots\sigma_m^k(\overline{2})$. Using Lemma 3.5 and the induction hypothesis, the resulting difference is once again $m^{\binom{k}{2}}(m-1)m^{k-1}$.

We still have to take into account the contributions of $\sigma_m^k(0)$ and $\sigma_m^k(\overline{1})$ within v. The word v begins with $m-1-j$ blocks of length m^k followed by $\sigma_m^k(\overline{1})\sigma_m^k(0)$, and ends with $j-1$ blocks of length m^k. We have to count the number of 0's appearing before $\sigma_m^k(\overline{1})$ and the $\overline{k+1}$'s appearing after $\sigma_m^k(0)$. There are $(m-1-j)m^{k-1}$ such 0's and $(m-3-j)m^{k-1}$ such $\overline{k+1}$'s. By comparing with the blocks occurring in the corresponding position in u, we obtain the following difference

$$\left(\binom{\sigma_m^k(\overline{j+1})}{\overline{1}\,\overline{2}\cdots\overline{k}} - \binom{\sigma_m^k(\overline{1})}{\overline{1}\,\overline{2}\cdots\overline{k}}\right)(m-1-j)m^{k-1} + \left(\binom{\sigma_m^k(\bar{j})}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^k(0)}{0\overline{1}\cdots\overline{k}}\right)(j-1)m^{k-1}.$$

By induction hypothesis, we find that both terms in parentheses are equal to $-m^{\binom{k}{2}}$. Therefore, the difference is $-m^{\binom{k}{2}}(m-2)m^{k-1}$.

Combining the results from the three preceding discussions, we get a total difference of

$$2m^{\binom{k}{2}}(m-1)m^{k-1} - m^{\binom{k}{2}}(m-2)m^{k-1} = m^{\binom{k+1}{2}}$$

matching the expected result. □

Corollary 4.2. *Let $u, v \in \mathcal{A}_m^*$ be of the same length and $i \in \mathcal{A}_m$. Then,*

$$\binom{\sigma_m^k(u)}{i(i-1)\cdots(i-k)} - \binom{\sigma_m^k(v)}{i(i-1)\cdots(i-k)} = (|u|_i - |v|_i)\, m^{\binom{k}{2}}.$$

In particular, if $u \gamma_1 v$, then $\sigma_m^k(u) \gamma_{k+1} \sigma_m^k(v)$.

Proof. As in the proof of the previous result, by Lemma 3.5, we may assume $i = 0$. There exist words p, u', and v' such that $u \sim_1 pu'$ and $v \sim_1 pv'$, where u' and v' share no common letters, and $|u'| = |v'|$. Let $\Psi(u) = (s_1, \ldots, s_m)$ and $\Psi(v) = (t_1, \ldots, t_m)$. Then, $\Psi(p) = (\min\{s_1, t_1\}, \ldots, \min\{s_m, t_m\})$. By Proposition 3.7, $\sigma_m^k(u) \sim_{k+1} \sigma_m^k(pu')$. Therefore,

$$\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{k}} = \binom{\sigma_m^k(pu')}{0\overline{1}\cdots\overline{k}} = \sum_{\substack{x,y\in\mathcal{A}_m^* \\ xy=0\overline{1}\cdots\overline{k}}} \binom{\sigma_m^k(p)}{x}\binom{\sigma_m^k(u')}{y}.$$

Thus,

$$\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^k(v)}{0\overline{1}\cdots\overline{k}} = \sum_{\substack{x,y\in\mathcal{A}_m^* \\ xy=0\overline{1}\cdots\overline{k}}} \binom{\sigma_m^k(p)}{x}\left(\binom{\sigma_m^k(u')}{y} - \binom{\sigma_m^k(v')}{y}\right).$$

Using Proposition 3.7 again, $\sigma_m^k(u') \sim_k \sigma_m^k(v')$. Therefore, if $|y| \leqslant k$, we have

$$\binom{\sigma_m^k(u')}{y} - \binom{\sigma_m^k(v')}{y} = 0.$$

Hence, we conclude

$$\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^k(v)}{0\overline{1}\cdots\overline{k}} = \binom{\sigma_m^k(u')}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^k(v')}{0\overline{1}\cdots\overline{k}}.$$

As shown in the proof of Proposition 4.1, since $\sigma_m^k(i) \sim_k \sigma_m^k(j)$ for all $i, j \in \mathcal{A}_m$, a non-zero difference arises only if a subword $0\overline{1}\cdots\overline{k}$ appears entirely within an m^k-block. More precisely, if $u' = a_1 \cdots a_r$ and $v' = b_1 \cdots b_r$ where a_i's and b_j's are letters, the difference can be expressed as

$$\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^k(v)}{0\overline{1}\cdots\overline{k}} = \sum_{i=1}^{r} \binom{\sigma_m^k(a_i)}{0\overline{1}\cdots\overline{k}} - \sum_{i=1}^{r} \binom{\sigma_m^k(b_i)}{0\overline{1}\cdots\overline{k}}.$$

Using Proposition 4.1 and recalling that u' and v' share no common letter, it follows that

$$\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^k(v)}{0\overline{1}\cdots\overline{k}} = (|u'|_0 - |v'|_0)\, m^{\binom{k}{2}}.$$

In the particular case where u and v are not abelian equivalent, the words u' and v' must be non-empty. Without loss of generality, we assume that 0 appears in u' (and does not appear in v'). Hence $(|u'|_0 - |v'|_0)\, m^{\binom{k}{2}} \neq 0$ and it follows that $\sigma_m^k(u)$ and $\sigma_m^k(v)$ do contain the same number of occurrences of the subword $0\overline{1}\cdots\overline{k}$. □

By combining Proposition 3.7 and Corollary 4.2, we obtain Proposition 2.3, which is restated below.

Proposition 2.3. *Let $x, y \in \mathcal{A}_m^*$ and $k \geqslant 1$. Then, $x \sim_1 y$ holds if and only if $\sigma_m^k(x) \sim_{k+1} \sigma_m^k(y)$.*

Proposition 4.1 dealt with subwords of length $k+1$ occurring in m^k-blocks. The next statement focuses on subwords of length at most k that appear in the image of a word under σ_m^k. This result will play a key role in the proof of Lemma 4.4.

Lemma 4.3. *Let $\ell \leqslant k$. For all j, the following holds*

$$\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{\ell-1}} = \binom{\sigma_m^k(u)}{\bar{j}\cdots\overline{j+\ell-1}}$$

Proof. Let $u = a_1 \cdots a_t$, where $a_i \in \mathcal{A}_m$. First of all, we note that trivially

$$\binom{\sigma_m^k(a_1 \cdots a_t)}{\bar{j}\cdots\overline{j+\ell-1}} = \binom{\tau_m^j(\sigma_m^k(a_1 \cdots a_t))}{\tau_m^j(\bar{j}\cdots\overline{j+\ell-1})},$$

as the subwords occur at the same positions in the respective words. Furthermore, we have $\tau_m^j(\bar{j}\cdots\overline{j+\ell-1}) = 0\overline{1}\cdots\overline{\ell-1}$. Finally, since $\sigma_m \circ \tau_m = \tau_m \circ \sigma_m$, it follows that

$$\binom{\sigma_m^k(a_1 \cdots a_t)}{\bar{j}\cdots\overline{j+\ell-1}} = \binom{\tau_m^j(\sigma_m^k(a_1 \cdots a_t))}{0\overline{1}\cdots\overline{\ell-1}} = \binom{\sigma_m^k(\tau_m^j(a_1 \cdots a_t))}{0\overline{1}\cdots\overline{\ell-1}} = \binom{\sigma_m^k((a_1+j)\cdots(a_t+j))}{0\overline{1}\cdots\overline{\ell-1}}.$$

Furthermore, by Proposition 3.7(iii), we know that $\sigma_m^k(a_1 \cdots a_t) \sim_k \sigma_m^k((a_1+j)\cdots(a_t+j))$. Hence, the desired result. □

The next lemma is presented in its full generality. For the sake of presentation, the proof is given in Section 10.

Lemma 4.4. *Let $k \geqslant 2$. Suppose $u, u', \gamma, \gamma', \delta, \delta' \in \mathcal{A}_m^*$ are words such that $\gamma\delta \sim_1 \gamma'\delta'$ and $|u| = |u'|$. Then, the difference*

$$\binom{\sigma_m^{k-1}(\gamma\, \sigma_m(u)\delta)}{0\overline{1}\cdots\overline{k}} - \binom{\sigma_m^{k-1}(\gamma'\, \sigma_m(u')\delta')}{0\overline{1}\cdots\overline{k}}$$

is given by

$$\begin{aligned} &m^{\binom{k}{2}}\Big[|u|_0 - |u'|_0 + |u|\left(|\gamma|_0 - |\gamma'|_0 + |\delta|_{\overline{1}} - |\delta'|_{\overline{1}}\right)\Big] \\ &\quad + m^{\binom{k}{2}-1} \sum_{b\in\mathcal{A}_m} \left(\binom{\gamma\delta}{b\overline{1}} - \binom{\gamma'\delta'}{b\overline{1}} + \binom{\gamma\delta}{0b} - \binom{\gamma'\delta'}{0b}\right). \end{aligned}$$

5 Recognizability and Structure of Factors

First, we recall a recognizability property stating that any long enough factor $U \in \mathrm{Fac}(\mathbf{t}_m)$ has a unique σ_m^k-factorization of the form $p_u\, \sigma_m^k(u) s_u$, where p_u and s_u are blocks of length less than m^k. Next, we examine the structure of those pairs (p_u, s_u) in detail and show that they are subject to strong constraints. This will allow us to carry out precise counting in Section 6.

We summarize some well-known concepts and results (see, for instance, [7, 11]). A morphism φ is called *marked* if, for every pair of distinct letters, their images under φ differ in both the first and last letters. A morphism $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ is said to be *primitive* if there exists an integer n such that, for all $a \in \mathcal{A}$, the word $\varphi^n(a)$ contains all letters of $\mathcal{A}$.

Remark 5.1. Let $\varphi : \mathcal{A}^* \to \mathcal{A}^*$ be a morphism, and let $n \geqslant 1$ be an integer. If φ is marked (respectively, primitive, ℓ-uniform), then φ^n has the same properties, meaning φ^n is marked (respectively, primitive, ℓ^n-uniform).

Note that, for all $k \geqslant 1$, the k^{th} power of our morphism of interest σ_m is such that $\sigma_m^k(i)$ begins with i and ends with $i-k$. Therefore, the morphism σ_m^k is marked.

Let $\mathbf{x}$ be a fixed point of a morphism φ over $\mathcal{A}$. A factor w of $\mathbf{x}$ is said to contain a *synchronization point* (w_1, w_2) if $w = w_1w_2$ and, for all $v_1, v_2 \in \mathcal{A}^*$, $s \in \mathrm{Fac}(\mathbf{x})$ such that $\varphi(s) = v_1w_1w_2v_2$, there exist $s_1, s_2 \in \mathrm{Fac}(\mathbf{x})$ such that $s = s_1s_2$, $\varphi(s_1) = v_1w_1$, and $\varphi(s_2) = w_2v_2$. A factor w that contains a synchronization point is said to be *circular*.

Proposition 5.2. *Let φ be an ℓ-uniform, primitive, marked morphism with $\mathbf{x}$ as one of its fixed points. If u is a circular factor of $\mathbf{x}$, then u has a unique φ-factorization (in the sense of Definition 2.1).*

Proposition 5.3 ([7, 11])**.** *For all $k \geqslant 1$, the morphism σ_m^k is an m^k-uniform, primitive, marked morphism. Moreover, every factor of its fixed point $\mathbf{t}_m$ that has length at least $2m^k$ is circular.*

Example 5.4. The factor $\sigma_m(0)^2$ of $\mathbf{t}_m$ has m factorizations $\sigma_m(00)$ and

$$\mathrm{suff}_j(\sigma_m(j)) \cdot \sigma_m(j) \cdot \mathrm{pref}_{m-j}(\sigma_m(j)), \qquad j = 1, \ldots, m-1.$$

However, only one of these is a σ_m-factorization, namely $\sigma_m(00)$. This is because 00 is a factor of $\mathbf{t}_m$ but j^3 does not occur in $\mathbf{t}_m$ for any j (cf. Proposition 3.3), implying that none of the other factorizations are indeed σ_m-factorizations.

The factor $\sigma_m(0)01\cdots(m-2)$ which has a length of $2m-1$, has two possible σ_m-factorizations:

$$\sigma_m(0) \cdot \mathrm{pref}_{m-1}(\sigma_m(0)) \quad \text{and} \quad \mathrm{suff}_{m-1}(\sigma_m(m-1)) \cdot \sigma_m(m-1).$$

Recall from Lemma 3.6 that 00 and $(m-1)(m-1)$ are indeed factors of $\mathbf{t}_m$.

Remark 5.5. For any $k \geqslant 1$, it is obvious that all factors of length at least $m^k - 1$ in $\mathbf{t}_m$ have a σ_m^k-factorization, since the image of a letter has length m^k. To simplify the arguments in Section 7, we extend this observation to all factors. Namely, for any $k \geqslant 1$, any factor U of $\mathbf{t}_m$ has a σ_m^k-factorization. We will prove this by induction on k.

For $k = 1$, the only case to consider is when a factor U appears properly within the image of a letter, i.e., $U = \ell \cdots (\ell + |U| - 1)$ for some $\ell \in \mathcal{A}_m$ with $|U| \leqslant m - 2$. Notice that

$$\mathrm{pref}_j(U) = \mathrm{suff}_j(\sigma_m(\ell + j)) \quad \text{and} \quad \mathrm{suff}_{|U|-j}(U) = \mathrm{pref}_{|U|-j}(\sigma_m(\ell + j)).$$

Since all squares a^2, where $a \in \mathcal{A}_m$, appear in $\mathbf{t}_m$, it follows that for each value of j, where $0 \leqslant j \leqslant |U|$, the word U has $|U| + 1$ distinct σ_m-factorizations of the form

$$\mathrm{suff}_j(\ell + j) \cdot \sigma_m(\varepsilon) \cdot \mathrm{pref}_{|U|-j}(\sigma_m(\ell + j)).$$

Now, assume that U has a σ_m^k-factorization of the form $x\sigma_m^k(u)y$, where x is a proper suffix of $\sigma_m^k(a)$ and y is a proper prefix of $\sigma_m^k(b)$, and aub is a factor of $\mathbf{t}_m$. If $u = \varepsilon$, then we have the σ_m^{k+1}-factorization $x \cdot \sigma_m^{k+1}(\varepsilon) \cdot y$. This is valid since $(a+1)b$ is a factor of $\mathbf{t}_m$, $\sigma_m^k(a)$ is a suffix of $\sigma_m^{k+1}(a+1)$, and $\sigma_m^k(b)$ is a prefix of $\sigma_m^{k+1}(b)$. Now, assume $|u| \geqslant 1$, implying $|U| \geqslant m^k$. If U does not appear properly within the σ_m^{k+1}-image of a letter, there is nothing to prove. Thus consider the case that U appears, without loss of generality, properly within $\sigma_m^{k+1}(0) = \sigma_m^k(0 \cdots (m-1))$, which implies $|U| \leqslant m^{k+1} - 2$. We can express U as $U = x'\sigma_m^k(u')y'$, where $u' = \ell(\ell+1)\cdots(\ell+t)$ for some $\ell \geqslant 1$ and $t < m-1-\ell$, with x' being a proper suffix of $\sigma_m^k(\ell-1)$, and y' a proper prefix of $\sigma_m^k(\ell + t + 1)$. Here, we allow $t = -1$ to indicate that u' is empty. For instance, we obtain the σ_m^{k+1}-factorization $x' \cdot \sigma_m^{k+1}(\varepsilon) \cdot \sigma_m^k(u')y'$, where x', being a suffix of $\sigma_m^k(\ell-1)$, is a proper suffix of $\sigma_m^{k+1}(\ell)$, and $u'y'$ is a proper prefix of $\sigma_m^{k+1}(\ell)$. As $\ell\ell$ is a factor of $\mathbf{t}_m$, the conclusion holds. If $x' = \varepsilon$, then we obtain the σ_m^{k+1}-factorization $\varepsilon \cdot \sigma_m^{k+1}(\varepsilon) \cdot \sigma_m^k(u')y'$. This concludes the proof.

Corollary 5.6. *For all factors $U \in \mathrm{Fac}(\mathbf{t}_m)$ of length $|U| \geqslant 2m^k$, there exists a unique σ_m^k-factorization:*

$$U = p_u\, \sigma_m^k(u) s_u.$$

In particular, the words p_u, s_u, and u are unique.

Proof. This result follows directly from Propositions 5.2 and 5.3. □

Example 5.7. Let $m = 3$ and $k = 2$. The word

$$U = 1200121202011202010122010121,$$

which has length 28, is a factor of $\mathbf{t}_3$. It can be factorized as:

$$\sigma_3(1)\sigma_3^2(01)\sigma_3(20)1$$

where $p_u = \sigma_3(1)$ and $s_u = \sigma_3(20)1$.

Since the word s_u is a proper prefix of some $\sigma_m^k(j)$, it has a specific structure. Since $|s_u| < m^k$, this length can be uniquely expressed using a base-m expansion as

$$|s_u| = \sum_{i=0}^{k-1} c_{k-i}\, m^i, \quad c_1, \ldots, c_k \in \{0, \ldots, m-1\}.$$

By applying a similar greedy procedure to the word s_u (refer to [10] for details on Dumont–Thomas numeration systems associated with a morphism, or [23]), we obtain the following unique decomposition

$$s_u = \prod_{i=1}^{k} \sigma_m^{k-i}(v_i) \tag{1}$$

where the words v_i are defined as follows

$$v_i = (j + \sum_{r=1}^{i-1} c_r)\,(j + \sum_{r=1}^{i-1} c_r + 1) \cdots (j + \sum_{r=1}^{i} c_r - 1).$$

Notice that $|v_i| = c_i$, and $v_1 \cdots v_k$ is a prefix of $(j(j+1)\cdots(j+m-1))^\omega$.

Example 5.8. The base-4 expansion of 226 is $3.4^3 + 2.4^2 + 2$. The prefix of $\sigma_4^4(0)$ with a length 226 is given by

$$\sigma_4^3(012)\sigma_4^2(30)12$$

where $\nu_1 = 012$, $\nu_2 = 30$, $\nu_3 = \varepsilon$, and $\nu_4 = 12$. Thus, $\nu_1 \cdots \nu_4 = 0123012$. For instance, $\sigma_4^3(\underline{02})\sigma_4^2(30)12$ is not the prefix of any $\sigma_4^4(a)$, as it involves applying σ_4^3 to a block composed of non-consecutive letters.

Remark 5.9. Knowing the value of j and the length $|s_u|$ uniquely determines the decomposition given in (1). Equivalently, for all $n \geqslant 1$ and letter a, there exists a unique factor of the form s_u, of length n, that starts (respectively, ends) with the letter a.

Corollary 5.10. *We collect the following facts.*

(i) With the above notation, let q (respectively, r) be the least (respectively, largest) integer such that c_q (or c_r) is non-zero. Let $\nu_q = xy$ and $\nu_r = zh$, such that $\nu_1 \cdots \nu_k = xy\nu_{q+1} \cdots \nu_{r-1}zh$. Then,

$$\sigma_m^{k-q}(y) \prod_{i=q+1}^{r-1} \sigma_m^{k-i}(\nu_i)\sigma_m^{k-r}(z)$$

is the proper prefix of the image of a letter under σ_m^k.

(ii) If $c_1 > 0$ and at least one of $c_2, \ldots, c_k$ is non-zero, the only admissible deletion of letters from ν_1, leading to a proper prefix of some $\sigma_m^k(a)$, is to suppress a prefix of ν_1. Removing a proper suffix of $\nu_1 = a_1 \cdots a_\ell$ or any "internal" factor $a_i \cdots a_j$ with $1 < i \leq j \leq \ell$ would violate the constraint that $\nu_1 \cdots \nu_k$ must be a prefix of the sequence $(j(j+1)\cdots(j+m-1))^\omega$.

(iii) If c_1 is the only non-zero coefficient, the only permissible deletion of letters from ν_1, resulting in a proper prefix of some $\sigma_m^k(a)$, is to suppress either a prefix or a suffix of ν_1.

A similar observation applies to p_u, which is the proper suffix of some $\sigma_m^k(j+1)$. The only difference lies in the fact that $\sigma_m(j+1)$ ends with j.

Since $|p_u| < m^k$, this length can be uniquely expressed using a base-m expansion as:

$$|p_u| = \sum_{i=0}^{k-1} c_{i+1}\, m^i, \quad c_k, \ldots, c_1 \in \{0, \ldots, m-1\}.$$

By applying a similar greedy procedure to the word p_u, we obtain the following decomposition:

$$p_u = \prod_{i=1}^{k} \sigma_m^{i-1}(\nu_i)$$

where the words ν_i are defined as follows

$$\nu_i = \left(j - k + i - \sum_{r=i}^{k} c_r + 1\right) \cdots \left(j - k + i - \sum_{r=i+1}^{k} c_r - 1\right)\left(j - k + i - \sum_{r=i+1}^{k} c_r\right).$$

Notice that $|\nu_i| = c_i$.

Example 5.11. The base-4 representation of 226 is $3 \cdot 4^3 + 2 \cdot 4^2 + 2$. Here, the suffix of $\sigma_4^4(0)$ with a length of 226 is given by

$$23\sigma_4^2(23)\sigma_4^3(123)$$

where $\nu_4 = 123$, $\nu_3 = 23$, $\nu_2 = \varepsilon$, and $\nu_1 = 23$.

Remark 5.12. Similar to the previous case, knowing the value of j and the length $|p_u|$ uniquely determines the decomposition. Equivalently, for all integers $n \geqslant 1$ and and any letter a, there exists a unique factor of the form p_u, of length n, that starts (respectively, ends) with the letter a.

Corollary 5.13. *We collect the following facts.*

(i) If $c_k > 0$ and at least one of $c_1, \ldots, c_{k-1}$ is non-zero, the only admissible deletion of letters from ν_k resulting in a proper suffix of some $\sigma_m^k(a)$ is to suppress a suffix of ν_k. Deleting a proper prefix of ν_k or some "internal" factor would not yield a valid suffix.

(ii) If c_k is the only non-zero coefficient, the only admissible deletion of letters from ν_k leading to a proper suffix of some $\sigma_m^k(a)$, is to suppress either a prefix or a suffix of ν_k.

6 Counting Classes of a New Equivalence Relation

Since σ_m is Parikh-constant, determining k-binomial equivalence of two factors primarily depends on their short prefixes and suffixes, rather than their central part composed of m^k-blocks. Thus, it is meaningful to focus on these prefixes and suffixes for our analysis. This section presents the core of our counting methods.

For the sake of presentation, let us recall Definition 2.2. Let $(p_1, s_1), (p_2, s_2) \in \mathcal{A}_m^{\leq m^k} \times \mathcal{A}_m^{\leq m^k}$. We have $(p_1, s_1) \equiv_k (p_2, s_2)$ whenever there exist $x, y, p, q, r, t \in \mathcal{A}_m^*$ with $|x|, |y| < m^{k-1}$ such that

$$\begin{aligned}(p_1, s_1) &= \big(x\sigma_m^{k-1}(p), \sigma_m^{k-1}(q)y\big),\\ (p_2, s_2) &= \big(x\sigma_m^{k-1}(r), \sigma_m^{k-1}(t)y\big),\end{aligned}$$

and one of the following conditions holds

- $pq \sim_1 rt$,
- $pq \sim_1 rt\,\sigma_m(0)$,
- $pq\,\sigma_m(0) \sim_1 rt$.

Notice that if $(p_1, s_1) \equiv_k (p_2, s_2)$, then

$$|p_1 s_1| = |p_2 s_2| \quad \text{or} \quad ||p_1 s_1| - |p_2 s_2|| = m^k.$$

Proposition 6.1. *Let $k \geqslant 2$, and $U, V \in \mathrm{Fac}(\mathbf{t}_m)$ of length at least $2m^k$. If $(p_u, s_u) \equiv_k (p_v, s_v)$, then $U \sim_k V$.*

Proof. Suppose first that $|p_u s_u| = |p_v s_v|$. By definition, there exist $x, y, p, q, r, t, u, v \in A^*$ such that:

$$\begin{aligned}U = p_u\,\sigma_m^k(u)s_u &= x\sigma_m^{k-1}(p)\sigma_m^k(u)\sigma_m^{k-1}(q)y &&= x\sigma_m^{k-1}(p\;\sigma_m(u)q)y\\ V = p_v\,\sigma_m^k(v)s_v &= x\sigma_m^{k-1}(r)\sigma_m^k(v)\sigma_m^{k-1}(t)y &&= x\sigma_m^{k-1}(r\;\sigma_m(v)t)y\end{aligned}$$

and $pq \sim_1 rt$. Since $|U| = |V|$, it follows that $|u| = |v|$ and $\sigma_m(u) \sim_1 \sigma_m(v)$. Thus, $p\,\sigma_m(u)q \sim_1 r\,\sigma_m(v)t$. By Proposition 3.7, we have

$$\sigma_m^{k-1}(p\;\sigma_m(u)q) \sim_k \sigma_m^{k-1}(r\;\sigma_m(v)t).$$

For the second case, suppose that $|p_u s_u| = |p_v s_v| + m^k$. Using the same notation as above, we have $pq \sim_1 rt\,\sigma_m(0)$ and $|v| = |u| + 1$. Therefore

$$r\,\sigma_m(v)t \sim_1 r\,\sigma_m(u)\,\sigma_m(0)t \sim_1 p\,\sigma_m(u)q$$

and we reach the same conclusion. □

We have an immediate lower bound for the k-binomial complexity of the generalized Thue–Morse word $\mathbf{t}_m$.

Corollary 6.2. *For all $n \geqslant 2m^k$, the k-binomial complexity $\mathsf{b}^{(k)}_{\mathbf{t}_m}(n)$ satisfies the inequality*

$$\mathsf{b}^{(k)}_{\mathbf{t}_m}(n) \geqslant \#\big(\{(p_u, s_u) \mid u \in \mathrm{Fac}_n(\mathbf{t}_m)\}/\equiv_k\big).$$

Later on, we will get the value of $\#(\{(p_u, s_u) \mid u \in \mathrm{Fac}_n(\mathbf{t}_m)\}/\equiv_k)$ with Theorem 6.5.

Let $n \geqslant 2m^k$. Define $n = \lambda \pmod{m^k}$, where $\lambda = \mu + \nu m^{k-1}$, with $\mu < m^{k-1}$ and $\nu < m$. We begin by defining a partition of the set of pairs.

Definition 6.3. Let $\ell \in \{0, \ldots, m^{k-1}\}$. Let

$$P^{(n)}_\ell := \big\{(p_u, s_u) \mid u \in \mathrm{Fac}_n(\mathbf{t}_m),\ |p_u| \equiv \ell \pmod{m^{k-1}}\big\}.$$

and similarly,

$$S^{(n)}_{\ell'} := \big\{(p_u, s_u) \mid u \in \mathrm{Fac}_n(\mathbf{t}_m),\ |s_u| \equiv \ell' \pmod{m^{k-1}}\big\}.$$

Note that

$$\bigcup_{\ell=0}^{m-1} P^{(n)}_\ell = \{(p_u, s_u) \mid u \in \mathrm{Fac}_n(\mathbf{t}_m)\} = \bigcup_{\ell'=0}^{m-1} S^{(n)}_{\ell'}.$$

Let $(p, s) \in P^{(n)}_\ell$. By Euclidean division, since $|p|, |s| < m^k$, we have

$$|p| = \ell + \alpha\, m^{k-1} \text{ and } \quad |s| = \ell' + \alpha'\, m^{k-1},$$

for some $\alpha, \alpha' < m$ and $\ell' < m^{k-1}$. We show that n, ℓ, α completely determine ℓ' and α'. In particular, for each ℓ, there exists a unique ℓ' such that $P^{(n)}_\ell = S^{(n)}_{\ell'}$.

Since

$$\ell + \ell' + (\alpha + \alpha')\, m^{k-1} = |ps| \equiv n \pmod{m^k},$$

we have

$$\ell + \ell' \equiv \mu \bmod m^{k-1}.$$

Thus, either

1. $\ell \leqslant \mu$ and $\ell' = \mu - \ell$ or,
2. $\ell > \mu$ and $\ell' = m^{k-1} + \mu - \ell$.

If $\ell \leqslant \mu$, then $\ell + \ell' = \mu$ and:

$$|ps| = \mu + (\alpha + \alpha')\, m^{k-1} \equiv \mu + \nu m^{k-1} \pmod{m^k}.$$

If $\alpha \leqslant \nu$, then $\alpha' = \nu - \alpha$. Otherwise $\alpha > \nu$, then $\alpha' = \nu + m - \alpha$.

In the second case ($\ell > \mu$), we have

$$\ell + \ell' = \mu + m^{k-1}, \quad \text{and} \quad |ps| = \mu + (\alpha + \alpha' + 1)\, m^{k-1}$$

If $\alpha \leqslant \nu - 1$, then $\alpha' = \nu - \alpha - 1$. Otherwise, $\alpha > \nu - 1$, then $\alpha' = \nu + m - \alpha - 1$. These observations are recorded in Table 2.

Example 6.4. Let $m = 3$ and $k = 2$. If $n \equiv 4 \pmod 9$, then $\mu = 1$ and $\nu = 1$. The set $P^{(n)}_0$ contains pairs (p, s) such that $|p| = 0, 3, 6$, which is $0 + \alpha 3$ for $\alpha = 0, 1, 2$. Since $\ell = 0 \leqslant 1 = \mu$, we have $\ell' = 1$. For $\alpha = 0$ or 1, which is less than or equal to ν, the corresponding values of α' are 1 and 0, respectively. For $\alpha = 2$ which is greater than ν, α' is $\nu + 3 - \alpha = 2$. Thus, the lengths of s corresponding to $|p| = 0, 3, 6$ are $4, 1, 7$, respectively. Therefore, note that $P^{(n)}_0 = S^{(n)}_1$.

The set $P^{(n)}_1$ contains pairs (p, s) such that $|p| = 1, 4, 7$, which is $1 + \alpha 3$ for $\alpha = 0, 1, 2$. Since $\ell = 1 \leqslant 1 = \mu$, we have $\ell' = 0$. For $\alpha = 0$ or 1, which is less than or equal to ν, the corresponding values of α' are 1 and 0, respectively. For $\alpha = 2$, which is greater than ν, α' is $\nu + 3 - \alpha = 2$.

	$\ell \leqslant \mu$		$\ell > \mu$
$\alpha \leqslant \nu$:	$\ell' = \mu - \ell$	$\alpha \leqslant \nu - 1$:	$\ell' = m^{k-1} + \mu - \ell$
	$\alpha' = \nu - \alpha$		$\alpha' = \nu - \alpha - 1$
i.e.,	$\alpha + \alpha' = \nu$	i.e.,	$\alpha + \alpha' = \nu - 1$
$\alpha > \nu$:	$\ell' = \mu - \ell$	$\alpha > \nu - 1$:	$\ell' = m^{k-1} + \mu - \ell$
	$\alpha' = \nu + m - \alpha$		$\alpha' = \nu + m - \alpha - 1$
i.e.,	$\alpha + \alpha' = \nu + m$	i.e.,	$\alpha + \alpha' = \nu + m - 1$

Table 2: Summary for (ℓ', α') for fixed μ, ν and α varying.

Thus, the lengths of s corresponding to $|p| = 1, 4, 7$ are $3, 0, 6$, respectively. Therefore, note that $P_1^{(n)} = S_0^{(n)}$.

The set $P_2^{(n)}$ contains pairs (p, s) such that $|p| = 2, 5, 8$, which is $2 + \alpha 3$ for $\alpha = 0, 1, 2$. Since $\ell = 2$ is greater than $\mu = 1$, we have $\ell' = 3 + \mu - 2 = 2$. For $\alpha = 0$, the corresponding value of α' is $\nu - 1 = 0$. For $\alpha = 1$ and $\alpha = 2$, both greater than $\nu - 1$, the corresponding values of α' are 2 and 1 respectively. Thus, the lengths of s corresponding to $|p| = 2, 5, 8$ are $2, 8, 5$, respectively. Finally, note that $P_2^{(n)} = S_2^{(n)}$.

Note that if $\mu = 0$, then $P_0^{(n)} = S_0^{(n)}$. If $\mu \neq 0$, then for $\ell = 0$, we have $\ell' = \mu \neq 0$. In that case $P_0^{(n)} = S_\mu^{(n)} \neq S_0^{(n)} = P_\mu^{(n)}$. This observation gives an initial hint as to why the statement of Theorem 6.5 contains two cases.

Recall that the abelian complexity of $\mathbf{t}_m$ is well known (see Theorem 1.7).

Theorem 6.5. *Let* $n \geqslant 2m^k$. *If* $\lambda = n \bmod m^k$ *and* $\lambda = \nu m^{k-1} + \mu$, *where* $\nu < m$ *and* $\mu < m^{k-1}$, *then the value of*

$$\#\{(p_u, s_u) \mid u \in \mathrm{Fac}_n(\mathbf{t}_m)\}/\equiv_k$$

is given by

$$(m^{k-1} - 1)(m^3 - m^2 + m) + \begin{cases} b_{\mathbf{t}_m}^{(k)}(m + \nu), & \text{if } \mu = 0; \\ m, & \text{otherwise.} \end{cases}$$

Remark 6.6. Note that for $k = 2$, which was the case studied in [18], this expression matches the 2-binomial complexity of $\mathbf{t}_m$. Thus, we obtain the converse of Proposition 6.1: Let u and v be two factors of $\mathbf{t}_m$ of length at least $2m^2$. Then, $u \sim_2 v$ if and only if $(p_u, s_u) \equiv_2 (p_v, s_v)$.

Proof. • **Case 1.a)** Let us consider $\mu \neq 0$ and $\ell \neq 0$. Assume that $\ell \leqslant \mu$. Referring to the first column of Table 2, the elements of $P_\ell^{(n)}$ have the form given in Table 3, where x^j and y^j are words and r_i^j, t_i^j are letters.

Since we are dealing with proper suffixes or prefixes of the image of a letter under σ_m^k, we also have

$$\forall j < m : \quad t_{i+1}^j = t_i^j - 1 \text{ and } r_{i+1}^j = r_i^j + 1.$$

Since $\ell \neq 0$ (respectively, $\ell \neq \mu$), the words x^j (respectively, y^j) are non-empty of length ℓ (respectively, $\mu - \ell$).

Thanks to Remarks 5.9 and 5.12, there are at most m^2 words on each row of Table 3: a prefix (respectively, suffix) of any given length is determined by its last (respectively, first) letter. Thanks to Lemma 3.6, there are exactly m^2 words on each row.

We now consider the quotient by $\equiv_k$. Since the words $r_1^0 \cdots r_{\nu-1}^0 r_\nu^0$ have length less than m and are made of consecutive letters, if two such words have distinct first letter, then there are not abelian equivalent. Hence the m^2 words on this row are pairwise non-equivalent.

The same argument applies on the second row. Nevertheless, if $t_1^1 = r_1^1 - 1$, then

$$(x^1 \sigma_m^{k-1}(t_1^1), \sigma_m^{k-1}(r_1^1 \cdots r_{\nu-1}^1) y^1) \equiv_k (x^1 \sigma_m^{k-1}(\varepsilon), \sigma_m^{k-1}(t_1^1 r_1^1 \cdots r_{\nu-1}^1) y^1).$$

α	p_u	s_u	α'
0	$x^0\sigma_m^{k-1}(\varepsilon)$	$\sigma_m^{k-1}(r_1^0\cdots r_{\nu-1}^0 r_\nu^0)y^0$	ν
1	$x^1\sigma_m^{k-1}(t_1^1)$	$\sigma_m^{k-1}(r_1^1\cdots r_{\nu-1}^1)y^1$	$\nu-1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\nu-1$	$x^{\nu-1}\sigma_m^{k-1}(t_{\nu-1}^{\nu-1}\cdots t_1^{\nu-1})$	$\sigma_m^{k-1}(r_1^{\nu-1})y^{\nu-1}$	1
ν	$x^\nu\sigma_m^{k-1}(t_\nu^\nu\, t_{\nu-1}^\nu\cdots\, t_1^\nu)$	$\sigma_m^{k-1}(\varepsilon)y^\nu$	0
$\nu+1$	$x^{\nu+1}\sigma_m^{k-1}(t_{\nu+1}^{\nu+1}t_\nu^{\nu+1}\cdots t_1^{\nu+1})$	$\sigma_m^{k-1}(r_1^{\nu+1}\cdots r_{\nu+1}^{\nu+1}\cdots r_{m-1}^{\nu+1})y^{\nu+1}$	$m-1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$m-1$	$x^{m-1}\sigma_m^{k-1}(t_{m-1}^{m-1}\cdots t_\nu^{m-1}\cdots t_1^{m-1})$	$\sigma_m^{k-1}(r_1^{m-1}\cdots r_{\nu+1}^{m-1})y^{m-1}$	$\nu+1$

Table 3: Words in $P_\ell^{(n)}$.

If $t_1^1 \neq r_1^1 - 1$, we cannot make such a move an keep equivalent pairs (we know from (1) that we must have consecutive letters in $t_1^1 r_1^1 \cdots r_{\nu-1}^1$). So we find $m(m-1)$ new classes.

We have a similar counting in the first $\nu+1$ rows (we proceed downwards, comparing elements on a row with elements on previous rows). Take a word of the form

$$x^j\sigma_m^{k-1}(t_j^j\cdots\, t_s^j\cdots\, t_1^j\,)$$

on the row $j \leqslant \nu$. Thanks to Corollary 5.10 (ii), we can only delete a suffix of $t_s^j\cdots\, t_1^j$ to keep a valid suffix of some $\sigma_m^k(a)$. If $t_1^j = r_1^j - 1$, since the suffix is made of consecutive letters

$$\begin{aligned}&(x^j\sigma_m^{k-1}(t_j^j\cdots t_s^j\cdots t_1^j), \sigma_m^{k-1}(r_1^j\cdots r_{\nu-r}^j)y^j)\\ \equiv_k\ &(x^j\sigma_m^{k-1}(t_j^j\cdots t_{s+1}^j), \sigma_m^{k-1}(t_s^j\cdots t_1^j r_1^j\cdots r_{\nu-r}^j)y^j)\end{aligned}$$

for any $1 \leqslant s \leqslant j$. We again find $m(m-1)$ new classes.

For the second part of the Table, take row $j \geqslant \nu+1$. The reasoning is again the same but this time, when $t_1^j = r_1^j - 1$, take $s \geqslant j-\nu$, then $t_s^j\cdots t_1^j\, r_1^j\cdots r_{m+\nu-j}^j$ has length $m+\nu-j+s \geqslant m$. So it has a prefix which is a cyclic permutation of $0, 1, \ldots, m-1$. Hence, so we find an equivalent pair

$$(x^j\sigma_m^{k-1}(t_j^j\cdots t_{s+1}^j), \sigma_m^{k-1}(r_{m-s}^j\cdots r_{m+\nu-j}^j)y^j)$$

in the first part of the table.

The case $\ell > \mu$ is treated similarly. As a conclusion, we have m^2 classes for the first row and $m(m-1)$ classes for each of the $m-1$ other rows for a total of $m^2 + m(m-1)^2$ classes.

We have considered so far $m^{k-1} - 2$ sets $P_\ell^{(n)}$ each containing $m^2 + m(m-1)^2$ classes.

• **Case 1.b)** Let us consider $\mu \neq 0$ and focus on $P_0^{(n)}$ (similar discussion for $P_\mu^{(n)}$). The only difference in Table 3 is that there is no word x^j (it is empty because $\ell = 0$). The word y^j remains non-empty (because $\mu \neq 0$). In the first row, we have $(\varepsilon, \sigma_m^{k-1}(r_1^0\cdots r_{\nu-1}^0 r_\nu^0)y^0)$ so the number of classes is given by the number m of choices for r_1^0. Now come the extra discussion for $1 \leqslant j \leqslant m-1$ due to the absence of x^j. In $\sigma_m^{k-1}(t_j^j\cdots\, t_1^j\,)$ to get equivalent pairs, we can as above move a suffix $t_s^j\cdots\, t_1^j$ to the second component whenever $t_1^j = r_1^j - 1$ but also move a prefix $t_j^j\cdots t_{j-s+1}^j$ whenever $t_{j-s+1}^j = r_1^j - 1$. Consequently, the word $t_j^j\cdots\, t_1^j$ should not contain $r_1^j - 1$ which is equivalent to $t_j^j \in \{r_1^j, r_1^j+1, \ldots, r_1^j+m-j-1\}$ using the fact that the word is made of consecutive letters. So we have $m(m-j)$ choices. So the total is given by

$$m + \sum_{j=1}^{m-1} m(m-j) = \frac{1}{2}\big(m^3 - m^2 + 2m\big),$$

and this contribution is doubled to take the symmetric case of $P_\mu^{(n)}$.

As a conclusion, when $\mu \neq 0$, i.e., if $n \neq 0 \pmod{m^k}$, then

$$\begin{aligned}\#\{(p_u, s_u) \mid u \in \mathrm{Fac}_n(\mathbf{t}_m)\}/\equiv_k &= (m^{k-1}-2)(m^2+m(m-1)^2)+m^3-m^2+2m \\ &= (m^{k-1}-1)(m^3-m^2+m)+m.\end{aligned}$$

• Case 2) Let $\mu = 0$. If $\ell \neq 0$, then from Table 2 we get $\ell' = m^{k-1}-\ell \neq 0$. Then, we have the same discussion as in our first case. The $m^{k-1}-1$ sets $P_\ell^{(n)}$ for $\ell = 1, \ldots, m-1$ contain $m^2+m(m-1)^2$ classes (we get the same main term in the expression).

If $\ell = 0$, then $\ell' = 0$. Here, the particularity of the single set $P_0^{(n)}$ is that in Table 3 the words x^j and y^j are both empty. So we only consider pairs (p_u, s_u) of the form $(\sigma_m^{k-1}(p'), \sigma_m^{k-1}(s'))$ with $|p'|, |s'| < m$ and $|p's'| = \nu$ or $m+\nu$. We will show that

$$\#(P_0^{(\nu\, m^{k-1})}/\equiv_k) = \#(\mathrm{Fac}_{2m+\nu}(\mathbf{t}_m)/\sim_1).$$

Thanks to Proposition 5.3, any factor x of length $2m+\nu$ has a unique factorization of the form

$$x = p_x\sigma(w)s_x \text{ with } |p_x|, |s_x| < m \text{ and } |w| \in \{1,2\}.$$

Thanks to Lemma 3.6, a pair $(p_u, s_u) = (\sigma_m^{k-1}(p'), \sigma_m^{k-1}(s'))$ belongs to $P_0^{(\nu\, m^{k-1})}$ if and only if (p', s') is of the form (p_x, s_x) for some x in $\mathrm{Fac}_{2m+\nu}(\mathbf{t}_m)$.

Let $x, y \in \mathrm{Fac}_{2m+\nu}(\mathbf{t}_m)$, and their corresponding factorizations $x = p_x\sigma(w)s_x$ and $y = p_y\sigma(w')s_y$. If $x \sim_1 y$ and $|p_xs_x| = |p_ys_y|$, then $|w| = |w'|$ and thus $\sigma_m(w) \sim_1 \sigma_m(w')$. So $p_xs_x \sim_1 p_ys_y$ and we get

$$(\sigma_m^{k-1}(p_x), \sigma_m^{k-1}(s_x)) \equiv_k (\sigma_m^{k-1}(p_y), \sigma_m^{k-1}(s_y)).$$

If $x \sim_1 y$ but $|p_xs_x| \neq |p_ys_y|$, then the difference of their length is m. We may assume that $|p_xs_x| = |p_ys_y| + m$, so $|w| = 1$ and $|w'| = 2$. Since $\sigma_m(a)$ is a circular permutation of $01\cdots(m-1)$, we deduce that $p_xs_x \sim_1 p_ys_y\sigma(0)$ and the same conclusion follows. The converse also holds, if $x, y \in \mathrm{Fac}_{2m+\nu}(\mathbf{t}_m)$ and $(\sigma_m^{k-1}(p_x), \sigma_m^{k-1}(s_x)) \equiv_k (\sigma_m^{k-1}(p_y), \sigma_m^{k-1}(s_y))$, then considering both situations, one concludes that $x \sim_1 y$. It is known that for words of length at least m the abelian complexity function is periodic of period m, see [8]. Hence,

$$\#(\mathrm{Fac}_{2m+\nu}(\mathbf{t}_m)/\sim_1) = \#(\mathrm{Fac}_{m+\nu}(\mathbf{t}_m)/\sim_1).$$

□

7 Characterizing Binomial Equivalence in $\mathbf{t}_m$

In this section, we focus on characterizing k-binomial equivalence among factors of $\mathbf{t}_m$ through their σ_m^{k-1}-factorizations. We recall the main result:

Proposition 2.4. *Let $k \geqslant 2$. For any two factors U and V of $\mathbf{t}_m$, the relation $U \sim_k V$ holds if and only if there exist σ_m^{k-1}-factorizations $U = p_u\,\sigma_m^{k-1}(u)s_u$ and $V = p_v\,\sigma_m^{k-1}(v)s_v$, such that $p_u = p_v$, $s_u = s_v$, and $u \sim_1 v$.*

We observe that this proposition extends [18, Thm. 2] by removing an additional assumption $|u|, |v| \geqslant 3$ and extending it to all $k \geqslant 2$.

To prove Proposition 2.4, we shall present the following restricted version.

Lemma 7.1. *Let $k \geqslant 2$ and U and V be factors of $\mathbf{t}_m$ for some $m \geqslant 2$. Assume further that U and V begin and end with distinct letters. Then $U \sim_k V$ if and only if there exist σ_m^{k-1}-factorizations $U = \sigma_m^{k-1}(u)$ and $V = \sigma_m^{k-1}(v)$ such that $u \sim_1 v$.*

Before diving into the proof of Lemma 7.1, let us observe how Proposition 2.4 follows from it. First, we obtain Theorem 1.12 as an immediate corollary of Lemma 7.1.

Theorem 1.12. *The length of the shortest factor which is k-binomially equivalent to a distinct factor is $2m^{k-1}$. In particular, for any length $n < 2m^{k-1}$, the k-binomial complexity $\mathfrak{b}^{(k)}_{\mathbf{t}_m}(n)$ coincides with the factor complexity $p_{t_m}(n)$.*

Proof. A pair of shortest distinct k-binomially equivalent factors necessarily begin and end with different letters due to k-binomial equivalence being cancellative (cf. Lemma 3.2). Lemma 7.1 thus shows that the pair of factors can be written in the form $\sigma_m^{k-1}(u)$ and $\sigma_m^{k-1}(v)$ with $u \sim_1 v$. Therefore, $|u| = |v| \geqslant 2$ (since they must begin and end with different letters), giving the lower bound. The pair $\sigma_m^{k-1}(01)$ and $\sigma_m^{k-1}(10)$, for example, gives the desired pair of length $2m^{k-1}$. □

We can now prove Proposition 2.4

Proof of Proposition 2.4. Let $k \geqslant 2$ be arbitrary. If U and V have the σ_m^{k-1}-factorizations $U = p_u\,\sigma_m^{k-1}(u)s_u$ and $V = p_v\,\sigma_m^{k-1}(v)s_v$, where $p_u = p_v$, $s_u = s_v$, and $u \sim_1 v$, then $U \sim_k V$ follows by Proposition 2.3 and the fact that $\sim_k$ is a congruence.

For the converse, assume $U \sim_k V$. There is nothing to prove if $U = V$, as all factors have a σ_m^{k-1}-factorization by Remark 5.5. So assume $U \neq V$. Write $U = pU's$ and $V = pV's$, where U' and V' begin and end with distinct letters. By cancellativity (Lemma 3.2), we have $U' \sim_k V'$. By Lemma 7.1, there exist σ_m^{k-1}-factorizations $U' = \sigma_m^{k-1}(u')$ and $V' = \sigma_m^{k-1}(v')$, where $u' \sim_1 v'$. Note that Theorem 1.12 implies $|U'|, |V'| \geqslant 2m^{k-1}$. By Corollary 5.6, these σ_m^{k-1}-factorizations are unique. It follows that U and V have the desired (unique) σ_m^{k-1}-factorizations $U = p\sigma_m^{k-1}(u')s$ and $V = p\sigma_m^{k-1}(v')s$, where $u' \sim_1 v'$. □

The proof of Lemma 7.1 proceeds by induction on k. We divide the remainder of the section into two subsections: the base case $k = 2$, handled in the first subsection, and the induction step, covered in the second. We observe that the base case $k = 2$ is almost handled by [18, Thm. 2], except that the additional assumption $|u|, |v| \geqslant 3$ appearing there needs to be removed. Although the cases where $|u|, |v| \leqslant 3$ could be treated separately, we provide a complete, independent, but similar, proof of the case $k = 2$, as it reveals our strategy for tackling the induction step.

7.1 Proof of Lemma 7.1 — The base case

We shall state the induction base case as a separate lemma:

Lemma 7.2. *Let U and V be factors of $\mathbf{t}_m$ that begin and end with distinct letters. Then $U \sim_2 V$ if and only if there exist σ_m-factorizations $U = \sigma_m(u)$ and $V = \sigma_m(v)$, such that $u \sim_1 v$.*

Proof. If such σ_m-factorizations exist for U and V, then the two words are 2-binomially equivalent by Proposition 2.3.

Assume that U and V are 2-binomially equivalent factors, beginning and ending with distinct letters. Let U and V have the σ_m-factorizations $p_u\,\sigma_m(u)s_u$ and $p_v\,\sigma_m(v)s_v$, respectively (such factorizations exist by Remark 5.5). Notice that $||u| - |v|| \leqslant 1$ due to length constraints. Without loss of generality, we assume that $|u| \leqslant |v|$. We have two cases to consider. Either the words have the same length or their lengths differ by one.

First, assume that $|u| = |v|$. If both s_u and s_v are empty, it follows that $p_u\,\sigma_m(u) \sim_1 p_v\,\sigma_m(v)$. Since $\sigma_m(u) \sim_1 \sigma_m(v)$, we conclude that $p_u \sim_1 p_v$. This further implies $p_u = \varepsilon = p_v$, as U and V start with distinct letters, and p_u and p_v are proper suffixes of images of letters. By Proposition 2.3, it follows that $u \sim_1 v$, thereby establishing the claimed factorizations.

Thus, we proceed under the assumption that at least one of the words s_u and s_v is non-empty, intending to get a contradiction. Without loss of generality, we assume that s_u is non-empty.

Now, let $\alpha-1$ denote the last letter of s_u. By assumption, we have $\binom{u}{\alpha(\alpha-1)} = \binom{v}{\alpha(\alpha-1)}$; applying Lemma 3.1 twice, we obtain

$$\begin{aligned}\binom{p_u s_u}{\alpha(\alpha-1)} + \binom{\sigma_m(u)}{\alpha(\alpha-1)} + |u|\Big(|p_u|_\alpha + |s_u|_{\alpha-1}\Big) \\ = \binom{p_v s_v}{\alpha(\alpha-1)} + \binom{\sigma_m(v)}{\alpha(\alpha-1)} + |v|\Big(|p_v|_\alpha + |s_v|_{\alpha-1}\Big).\end{aligned}$$

Observe that $|p_w|_\alpha = |p_w s_w|_\alpha - |s_w|_\alpha$, where w is either u or v. Similarly, we have $|s_u|_{\alpha-1} = 1$ and $|s_u|_\alpha = 0$. Substituting these values into the previous equation yields

$$\begin{aligned}\binom{p_u s_u}{\alpha(\alpha-1)} + \binom{\sigma_m(u)}{\alpha(\alpha-1)} + |u|(|p_u s_u|_\alpha + 1) \\ = \binom{p_v s_v}{\alpha(\alpha-1)} + \binom{\sigma_m(v)}{\alpha(\alpha-1)} + |v|\big(|p_v s_v|_\alpha - |s_v|_\alpha + |s_v|_{\alpha-1}\big).\end{aligned}$$

The terms $|u||p_u s_u|_\alpha$ and $|v||p_v s_v|_\alpha$ cancel because $|u| = |v|$, and the equivalence $U \sim_2 V$ implies $p_u s_u \sim_1 p_v s_v$. By Lemma 3.4, $\alpha(\alpha-1)$ appears exclusively in $\sigma_m(\alpha)$, implying that $\binom{\sigma_m(u)}{\alpha(\alpha-1)} = |u|_\alpha$. Rearranging this equation yields the following equality

$$|u|_\alpha + |v|(|s_v|_\alpha - |s_v|_{\alpha-1}) = |v|_\alpha - |u| + \binom{p_v s_v}{\alpha(\alpha-1)} - \binom{p_u s_u}{\alpha(\alpha-1)}. \tag{2}$$

Claim 1. *1) The left-hand side of* (2) *is non-negative. Furthermore, it is equal to* 0 *if and only if either* $u = v = \varepsilon$, *or* $|u|_\alpha = 0$ *and* $|s_v|_\alpha = |s_v|_{\alpha-1}$.

2) The right-hand side of (2) *is non-positive. Moreover, it equals* 0 *if and only if* $|v|_\alpha = |v|$ *and* α *does not appear in* $p_u s_u$.

Proof of claim: Consider the first claim. Note that the left-hand side can only be negative if $|s_v|_{\alpha-1} > |s_v|_\alpha$. However, this situation cannot occur: if $\alpha-1$ appears in s_v, then as s_v does not end with $\alpha-1$; instead, $\alpha-1$ must be followed by α. Consequently, the coefficient of $|v|$ is non-negative, showing the non-negativity of the left-hand side. To attain a value of 0, we must have that either $u = \varepsilon$, or $|u|_\alpha = 0$ and $|s_v|_\alpha = |s_v|_{\alpha-1}$.

Let us consider the second claim. If α does not appear in $p_u s_u$, then

$$\binom{p_v s_v}{\alpha(\alpha-1)} = 0 = \binom{p_u s_u}{\alpha(\alpha-1)}.$$

Consequently, the right-hand side is equal to $|v|_\alpha - |v|$, which is clearly non-positive, and it is equal to 0 if and only if $|v|_\alpha = |v|$.

If α appears in $p_u s_u$, it must occur in p_u and does so precisely once. Since $\alpha-1$ does not appear in p_u after α, we have $\binom{p_u s_u}{\alpha(\alpha-1)} = 1$. Next, consider the occurrences of $\alpha-1$ and α in $p_v s_v$. Note that α cannot precede $\alpha-1$ in p_v or s_v. If $\alpha-1$ appears in s_v then, because s_v does not end with $\alpha-1$, it must be followed by α in s_v. Thus, we conclude that $\binom{p_v s_v}{\alpha(\alpha-1)} = 0$. Hence, the right-hand side equals $|v|_\alpha - |v| - 1$, which is strictly negative. The desired conclusion thereby follows. ■

The above claim shows that (2) can only be satisfied when both the left-hand side and the right-hand side are equal to zero. In other words, α must not appear in $p_u s_u$ (and consequently not in $p_v s_v$) and either: (a) $u = v = \varepsilon$; or (b) $|u|_\alpha = 0$, $|s_v|_{\alpha-1} = |s_v|_\alpha = 0$, and $|v|_\alpha = |v|$. Note that p_v must contain $\alpha-1$, which corresponds to the occurrence of $\alpha-1$ as the last letter of s_u, and thus p_v must end with $\alpha-1$; otherwise, it would contain α immediately following $\alpha-1$. This situation is illustrated in Fig. 2. Since $|u|_\alpha = 0$, the image of each letter of u under σ_m contains the factor $\alpha(\alpha-1)$. Since $v = \alpha^{|v|}$, the image of each letter of v under σ_m begins with α and ends with $\alpha-1$.

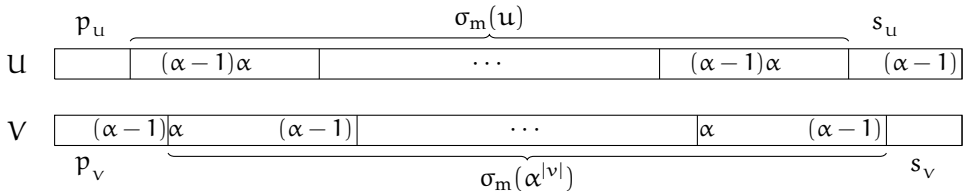


Figure 2: Illustrating the situation $|u| = |v|$ and s_u or s_v non-empty.

Consider the sum

$$\sum_{x \in \mathcal{A}_m} \binom{U}{(\alpha-1)x} + \binom{U}{x\alpha} - \binom{V}{(\alpha-1)x} - \binom{V}{x\alpha},$$

which equals zero, based on the assumption that $U \sim_2 V$. Observe that $\sum_{x \in \mathcal{A}_m} \binom{U}{(\alpha-1)x}$ counts, for each occurrence of $(\alpha-1)$ in U, the number of letters to its right. Similarly, $\sum_{x \in \mathcal{A}_m} \binom{U}{x\alpha}$ counts, for each occurrence of α in U, the number of letters to its left. With this interpretation, the "positive" part of the sum is equal to $|u| \cdot |U|$. Each of the $|u|$ occurrences of the factor $(\alpha-1)\alpha$ contributes $|U|$ to the positive count, while the last occurrence of $\alpha-1$ contributes zero. Similarly, the negative part of the sum is equal to $-|v| \cdot |V| - |s_v|$. Each of the $|v|$ occurrences of the factor $(\alpha-1)\alpha$ contributes $-|v| \cdot |V|$ to the negative count, while the last occurrence of $\alpha-1$ contributes $-|s_v|$. Since the sum must equal zero, we conclude that $s_v = \varepsilon$. However, now V ends with $\alpha-1$: if $v \neq \varepsilon$, then $\sigma_m(v)$ ends with $\alpha-1$, and if $v = \varepsilon$, then $p_v = V$ ends with $\alpha-1$. This contradicts the assumption that U and V end with distinct letters, resulting in a contradiction when $|u| = |v|$ and at least one of the words s_u and s_v is non-empty.

Second, assume that $|u| + 1 = |v|$. We will show that this case is impossible as it leads to a contradiction. In this situation, s_u must be non-empty (as must p_u), since $|p_u s_u| = |p_v s_v| + m$ and $|p_u|, |s_u| < m$. Let $\alpha-1$ be the last letter of s_u. Let β be the first letter of $v = \beta v'$, where $|v'| = |u|$, and let $p'_v = p_v\, \sigma_m(\beta)$. Note that $p_u s_u \sim_1 p'_v s_v$. As before, we have

$$\binom{U}{\alpha(\alpha-1)} = \binom{V}{\alpha(\alpha-1)}.$$

Using similar techniques as in the previous case, the equality can be expressed equivalently as

$$|u|_\alpha + |v'|(|s_v|_\alpha - |s_v|_{\alpha-1}) = |v'|_\alpha - |u| + \binom{p'_v s_v}{\alpha(\alpha-1)} - \binom{p_u s_u}{\alpha(\alpha-1)}.$$

We may proceed similarly as in the previous case. It is clear that the left-hand side is non-negative, and it equals zero if and only if either $u = v' = \varepsilon$ or $|s_v|_\alpha = |s_v|_{\alpha-1}$ and $|u|_\alpha = 0$.

Claim 2. *The right-hand side is non-positive and, moreover, equals zero if and only if $v' = \alpha^i$ and $\beta = \alpha$.*

Proof of claim: To begin, we show that $\binom{p_u s_u}{\alpha(\alpha-1)} = 1$. Since α appears in $p'_v s_v$ (in $\sigma_m(\beta)$), it must also appear in $p_u s_u$; since it does not appear in s_u, it appears in p_u. Furthermore, there is exactly one occurrence of α in $p_u s_u$. It should be noted that in p_u, $\alpha-1$ can precede α (if it appears at all) since $|p_u| < m$. Hence, there is only one occurrence of the subword $\alpha(\alpha-1)$, as desired.

Next, we consider $\binom{p'_v s_v}{\alpha(\alpha-1)}$. Observe that s_v does not contain $\alpha-1$; if it did, then it would be followed by a second occurrence of α in $p'_v s_v$ since it cannot end with $\alpha-1$, resulting in a contradiction.

Since α appears in $\sigma_m(\beta)$ within $p'_v s_v$ (and only once), we conclude that $\binom{p'_v s_v}{\alpha(\alpha-1)} = 1$ if and only if $\beta = \alpha$. Otherwise, $\binom{p'_v s_v}{\alpha(\alpha-1)} = 0$. Consequently, the right-hand side is non-positive and equals 0 if and only if $|v'| = |v'|_\alpha$ and $\beta = \alpha$. ■

For the equation above to be satisfied, we must have $|u|_\alpha = 0$, $v' = \alpha^i$ for some $i \geqslant 0$, and $\beta = \alpha$. Additionally, we have established that $|s_v|_\alpha = |s_v|_{\alpha-1} = 0$, regardless of whether $u = \varepsilon$ or not. It should be noted that if $\alpha - 1$ appears for a second time in $p_u s_u$, it must occur just before α in p_u and as the last letter of p_v; otherwise $p'_v s_v$ would contain a second occurrence of α. If $\alpha - 1$ appears only once in $p_u s_u$, then p_u begins with α. Fig. 3 illustrates the situation (the possible occurrences of $\alpha - 1$ in p_u and p_v are not shown).

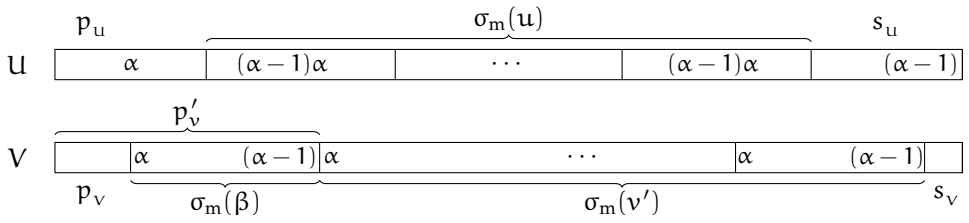


Figure 3: Illustrating the situation $|u| + 1 = |v|$.

Consider now the sum

$$\sum_{x \in \mathcal{A}_m} \binom{U}{(\alpha-1)x} + \binom{U}{x\alpha} - \binom{V}{(\alpha-1)x} - \binom{V}{x\alpha},$$

which is equal to 0 due to the assumption that $U \sim_2 V$. If $\alpha - 1$ does not appear in p_u (and thus not in p_v), then the positive side equals $|u||U|$; recall that p_u begins with α in this case. The negative side equals $-|v'||V| - |p_v s_v|$. This implies that $p_v s_v = \varepsilon$. But then $V = \sigma_m(\alpha^{i+1})$ ends with $\alpha - 1$, a contradiction.

If $\alpha - 1$ does appear in p_u and p_v, then the positive side is equal to $(|u| + 1)|U|$ whereas the negative side is equal to $-(|v'| + 1)|V| - |s_v|$. Hence, $s_v = \varepsilon$, and again V ends with $\alpha - 1$. This shows that the case where $|u| + 1 = |v|$ is impossible.

We have shown that the only possible way for $U \sim_2 V$ to hold is by having the claimed σ_m-factorizations, thus completing the proof. □

7.2 Proof of Lemma 7.1 — The induction step

Proof of Lemma 7.1. Suppose the two factors U and V possess the σ_m^{k-1}-factorizations $U = \sigma_m^{k-1}(u)$ and $V = \sigma_m^{k-1}(v)$, where $u \sim_1 v$. In that case, they are k-binomially equivalent, as stated in Proposition 3.7.

We consider the converse claim by induction on k, starting with the base case $k = 2$ which is addressed by Lemma 7.2. Assume that the claim holds for some $k \geqslant 2$, and consider $U \sim_{k+1} V$ with $U \neq V$, beginning and ending with distinct letters. Suppose U and V have σ_m^k-factorizations of the form $p_u \sigma_m^k(u) s_u$ and $p_v \sigma_m^k(v) s_v$, respectively, where $|u|, |v| \geqslant 0$ (note that such factorizations are guaranteed by Remark 5.5). By factoring out full σ_m^{k-1}-images from p_u, s_u, p_v, and s_v, we obtain the corresponding σ_m^{k-1}-factorizations of the form

$$U = p'_u \sigma_m^{k-1}(\gamma_u \, \sigma_m(u) \delta_u) s'_u \quad \text{and} \quad V = p'_v \sigma_m^{k-1}(\gamma_v \, \sigma_m(u) \delta_v) s'_v,$$

where $p_w = p'_w \sigma_m^{k-1}(\gamma_w)$ and $s_w = \sigma_m^{k-1}(\delta_w) s'_w$ for $w \in \{u, v\}$. Under this assumption, it follows that $U \sim_k V$, and by the induction hypothesis, we have $p'_w s'_w = \varepsilon$ for $w \in \{u, v\}$. Furthermore, $\gamma_u \, \sigma_m(u) \delta_u \sim_1 \gamma_v \, \sigma_m(v) \delta_v$, where the words U and V begin and end with distinct letters.

First, assume that $|u| = |v|$. Then $\gamma_u \delta_u \sim_1 \gamma_v \delta_v$. If both δ_u and δ_v are empty, it follows that $\gamma_u \sim_1 \gamma_v$. Since γ_u and γ_v are suffixes of σ_m-images of letters, they must be equal. Moreover, since U and V begin with distinct letters, this implies that $\gamma_u = \gamma_v = \varepsilon$. Thus, we have $U = \sigma_m^k(u)$ and $V = \sigma_m^k(v)$, confirming the claimed σ_m^k-factorizations by Proposition 2.3.

We now proceed to the case where either δ_u or δ_v is non-empty. Without loss of generality, we may assume that that $\delta_u \neq \varepsilon$, and let $\alpha - 1$ denote its final letter. In particular, α does not occur in

δ_u. We can apply Lemma 4.4 to $\sigma_m^{k-1}(\gamma_u\,\sigma_m(u)\delta_u)$ and $\sigma_m^{k-1}(\gamma_v\,\sigma_m(v)\delta_v)$, using $\alpha(\alpha-1)\cdots$ in place of $0\overline{1}\cdots$. Since these two words are assumed to be $(k+1)$-binomially equivalent, we obtain, by dividing by $m^{\binom{k}{2}-1}$

$$0 = m\Big[|u|_\alpha - |v|_\alpha + |u|\,(|\gamma_u|_\alpha - |\gamma_v|_\alpha + |\delta_u|_{\alpha-1} - |\delta_v|_{\alpha-1})\Big] \\ + \sum_{b\in\mathcal{A}_m}\left(\binom{\gamma_u\delta_u}{b(\alpha-1)} - \binom{\gamma_v\delta_v}{b(\alpha-1)} + \binom{\gamma_u\delta_u}{\alpha b} - \binom{\gamma_v\delta_v}{\alpha b}\right).$$

By observing that $|\gamma_w|_\alpha = |\gamma_w\delta_w|_\alpha - |\delta_w|_\alpha$, where $w \in \{u, v\}$, we can simplify the first term as follows

$$\begin{aligned} &m\big[|u|_\alpha - |v|_\alpha + |u|(|\delta_u|_{\alpha-1} - |\delta_u|_\alpha - |\delta_v|_{\alpha-1} + |\delta_v|_\alpha)\big] \\ &\quad = m\big[|u|_\alpha + |u|(|\delta_v|_\alpha - |\delta_v|_{\alpha-1})\big] + m(|v| - |v|_\alpha). \end{aligned}$$

Let us define Δ_α as

$$\Delta_\alpha := \sum_{b\in\mathcal{A}_m}\left(\binom{\gamma_v\delta_v}{b(\alpha-1)} + \binom{\gamma_v\delta_v}{\alpha b} - \binom{\gamma_u\delta_u}{b(\alpha-1)} - \binom{\gamma_u\delta_u}{\alpha b}\right).$$

Rearranging the previous equation, we obtain

$$m\big[|u|_\alpha + |u|(|\delta_v|_\alpha - |\delta_v|_{\alpha-1})\big] = m(|v|_\alpha - |v|) + \Delta_\alpha. \tag{3}$$

Recall that $|\delta_u|_{\alpha-1} = 1$ and $|\delta_u|_\alpha = 0$. Notice that the left-hand side is non-negative, and the only way where it could become negative is if $\alpha-1$ appears in δ_v. However, since δ_v does not end with $\alpha-1$ (as δ_u ends with it), this occurrence of $\alpha-1$ must be followed by α. Furthermore, the left-hand side is equal to zero if and only if $|u|_\alpha = 0$ and $|\delta_v|_\alpha = |\delta_v|_{\alpha-1}$.

Next, we show that the right-hand side is non-positive. Indeed, since $m(|v|_\alpha - |v|)$ is non-positive, it is sufficient to show that the sum Δ_α is also non-positive.

Claim 3. *The value of Δ_α is $-|\delta_v|$ if and only if $|\gamma_u\delta_u|_{\alpha-1} = |\gamma_u\delta_u|_\alpha + 1$. Otherwise, $\Delta_\alpha = -|\gamma_u\delta_u|$. Moreover, in the former case, $\alpha-1$ is the last letter of γ_v.*

Proof of claim: We first observe that

$$\sum_{b\in\mathcal{A}_m}\binom{x}{\alpha b}$$

counts, for each occurrence of the letter α in the word x, the number of letters that occur to its right. Similarly,

$$\sum_{b\in\mathcal{A}_m}\binom{x}{b(\alpha-1)},$$

counts, for each occurrence of $\alpha-1$, the number of letters occurring to its left.

We then consider the occurrences of α and $\alpha-1$ in the two words $\gamma_u\delta_u$ and $\gamma_v\delta_v$, as well as their contributions to the sum Δ_α. Notice that since δ_u does not contain α, there is at most one occurrence of α in $\gamma_u\delta_u$. Furthermore, there can be at most two occurrences of $\alpha-1$.

First of all, note that α does not appear in δ_u. The contribution from $\alpha-1$, as the last letter of δ_u, to the term $-\binom{\gamma_u\delta_u}{b(\alpha-1)}$ results in a value of $-|\gamma_u\delta_u|+1$ to Δ_α.

- First, assume that $|\gamma_u\delta_u|_{\alpha-1} = 1$. We proceed by dividing this into additional subcases, considering whether $\alpha-1$ appears in δ_v or not.
  - If δ_v contains $\alpha-1$, then this occurrence must be followed by by α. These two occurrences provide $|\gamma_v\delta_v|-2$ towards Δ_α. Now, α must appear in γ_u, whereas $\alpha-1$ should not. This situation occurs only if γ_u starts with α, as it is a suffix of the image of a letter (as depicted in Fig. 4).

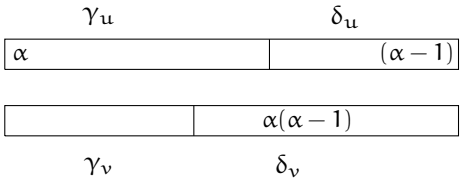


Figure 4: Illustrating the situation $|\gamma_u\delta_u|_{\alpha-1} = |\gamma_v|_{\alpha-1} = 1$.

This occurrence contributes $-|\gamma_u\delta_u| + 1$ towards Δ_α. Consequently, in this case, we have:

$$\Delta_\alpha = -|\gamma_u\delta_u| + 1 + |\gamma_v\delta_v| - 2 - |\gamma_u\delta_u| + 1 = -|\gamma_u\delta_u|.$$

Moreover, in this situation, we also have

$$|\gamma_u\delta_u|_\alpha = |\gamma_u\delta_u|_{\alpha-1}.$$

– If δ_v does not contain $\alpha - 1$, then γ_v contains $\alpha - 1$. We then further split this case based on whether α appears in $\gamma_u\delta_u$ or not.

  * Assume that α appears in $\gamma_v\delta_v$. In this case, either γ_v contains α as the letter directly following $\alpha - 1$ with $|\delta_v|_\alpha = 0$, or $\alpha - 1$ is the last letter of γ_v and δ_v begins with α (because, in this case, $\alpha-1$ does not appear in δ_v). In both cases, we have $\alpha - 1$ followed by α in $\gamma_v\delta_v$, resulting in a contribution of $|\gamma_v\delta_v| - 2$. Now, α appears in γ_u, while $\alpha - 1$ does not. This is possible only when α is the first letter of γ_u, thus contributing $-|\gamma_u\delta_u| + 1$ towards Δ_α. Hence, we find

    $$\Delta_\alpha = -|\gamma_u\delta_u| + 1 + |\gamma_v\delta_v| - 2 - |\gamma_u\delta_u| + 1 = -|\gamma_u\delta_u|.$$

    Note also that in this case

    $$|\gamma_u\delta_u|_\alpha = |\gamma_u\delta_u|_{\alpha-1}.$$

  * Assume that α does not occur in $\gamma_v\delta_v$. Consequently, the occurrence $\alpha - 1$ in γ_v must be its last letter. Therefore, in this case, we have

    $$\Delta_\alpha = -|\gamma_u\delta_u| + 1 + |\gamma_v| - 1 = -|\delta_v|.$$

    Moreover, in this case, we have

    $$|\gamma_u\delta_u|_\alpha + 1 = |\gamma_u\delta_u|_{\alpha-1}$$

    and $\alpha - 1$ is the last letter of γ_u.

- Assume secondly that$|\gamma_u\delta_u|_{\alpha-1} = 2$. Therefore, δ_v must contain $\alpha - 1$, and this is followed by α since δ_v cannot end with $\alpha-1$. These occurrences contribute $|\gamma_v\delta_v|-2$ to Δ_α. Now, γ_u also contains $\alpha-1$. Since α appears in δ_v, it must also occur in γ_u causing the two letters to appear consecutively. These occurrences contribute $-|\gamma_u\delta_u|+2$ to Δ_α. Finally, we consider the contribution of $\alpha - 1$ in γ_v. Since α is already present in δ_v it cannot occur in γ_v, thus γ_v ends with $\alpha - 1$. This provides Δ_α with $|\gamma_v| - 1$. Fig. 5 illustrates this situation.

  Thus, in this case, we have

  $$\begin{aligned}\Delta_\alpha &= -|\gamma_u\delta_u| + 1 + |\gamma_v\delta_v| - 2 - |\gamma_u\delta_u| + 2 + |\gamma_v| - 1 \\ &= -|\gamma_u\delta_u| + |\gamma_v| = -|\delta_v|.\end{aligned}$$

  Observe once more that

  $$|\gamma_u\delta_u|_\alpha + 1 = |\gamma_u\delta_u|_{\alpha-1},$$

  and furthermore, $\alpha - 1$ is the last letter of γ_v.

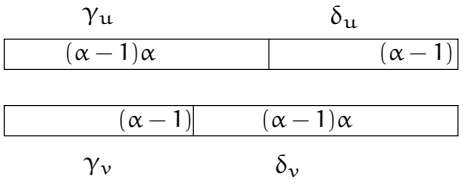


Figure 5: Illustrating the situation $|\gamma_u\delta_u|_{\alpha-1} = 2$.

All cases have been considered, and each one leads to the desired conclusion. ■

The preceding claim indicates that Δ_α in (3) is non-positive. For (3) to hold true, it must be the case that $\delta_v = \varepsilon$ and $|v|_\alpha = |v|$. Moreover, γ_v ends with $\alpha-1$ as stated in the above claim. However, $\gamma_v\,\sigma_m(v)$ ends with $\alpha-1$: either γ_v ends with $\alpha-1$ when $v \neq \varepsilon$, or γ_v ends with $\alpha-1$ if $v = \varepsilon$. This conclusion contradicts the initial assumption that the words end with distinct letters. Therefore, we have shown that the case $|u| = |v|$ is impossible if either of the words s_u or s_v is empty.

Second, assume that $|u| \neq |v|$**.** Due to the length constraints, it follows that $||u| - |v|| = 1$. Without loss of generality, let us assume that $|v| = |u| + 1$ and express v in the form $v = \beta v'$, where $\beta \in \mathcal{A}_m$. Consequently, we have $\gamma_u\delta_u \sim_1 \gamma_v\,\sigma_m(\beta)\delta_v$ implying that both γ_u and δ_u are non-empty. Let $\alpha-1$ denote the last letter of δ_u.

We may now apply Lemma 4.4, since we have $|u| = |v'|$ and $\gamma_u\delta_u \sim_1 \gamma_v\,\sigma_m(\beta)\delta_v$ (with α in place of 0). Rewriting γ'_v as $\gamma_v\,\sigma_m(\beta)$, we obtain (after dividing both sides by $m^{\binom{k}{2}-1}$)

$$0 = m\Big[|u|_\alpha - |v'|_\alpha + |u|\,(|\gamma_u|_\alpha - |\gamma'_v|_\alpha + |\delta_u|_{\alpha-1} - |\delta_v|_{\alpha-1})\Big] + \sum_{b\in\mathcal{A}_m}\left(\binom{\gamma_u\delta_u}{b(\alpha-1)} - \binom{\gamma'_v\delta_v}{b(\alpha-1)} + \binom{\gamma_u\delta_u}{\alpha b} - \binom{\gamma'_v\delta_v}{\alpha b}\right).$$

Write again

$$|\gamma_u|_\alpha - |\gamma'_v|_\alpha + |\delta_u|_{\alpha-1} - |\delta_v|_{\alpha-1} = |\delta_u|_{\alpha-1} - |\delta_u|_\alpha - |\delta_v|_{\alpha-1} + |\delta_v|_\alpha.$$

Furthermore, defining

$$\Delta_\alpha = \sum_{b\in\mathcal{A}_m}\left(\binom{\gamma'_v\delta_v}{b(\alpha-1)} + \binom{\gamma'_v\delta_v}{\alpha b} - \binom{\gamma_u\delta_u}{b(\alpha-1)} - \binom{\gamma_u\delta_u}{\alpha b}\right),$$

and recalling that $|\delta_u|_{\alpha-1} = 1$ and $|\delta_u|_\alpha = 0$, the preceding equation simplifies to

$$m(|u|_\alpha + |u|(|\delta_v|_\alpha - |\delta_v|_{\alpha-1})) = m(|v'|_\alpha - |v'|) + \Delta_\alpha. \tag{4}$$

Using arguments analogous to those in **Claim 1**, the left-hand side is shown to be non-negative. Moreover, it equals zero if and only if $|u|_\alpha = 0$ and $|\delta_v|_\alpha = |\delta_v|_{\alpha-1}$. Additionally, we compute the right-hand side in an analogous manner, showing that it is non-positive.

Claim 4. *We have* $\Delta_\alpha = -|\delta_v|$*, or* $\Delta_\alpha = -|\gamma_v\delta_v|$ *if and only if* $\beta = \alpha$*. In all other cases,* $\Delta_\alpha = -|\gamma_u\delta_u|$ *or* $\Delta_\alpha = -m - |\delta_v|$*.*

Proof of claim: We once again consider the occurrences of α and $\alpha-1$ in the two words $\gamma_u\delta_u$ and $\gamma'_v\delta_v$, and examine their contributions to the sum Δ_α.

Recall that $\alpha-1$ is the last letter of δ_u. Therefore, α can appear at most once in $\gamma_u\delta_u$. Since $\gamma_u\delta_u \sim_1 \gamma'_v\delta_v = \gamma_v\,\sigma_m(\beta)\delta_v$, and α appears in $\sigma_m(\beta)$, we conclude that α appears precisely once in $\gamma_u\delta_u$, and therefore must appear in γ_u.

• **Occurrences in** δ_u **and** $\sigma_m(\beta)$**:** The occurrence of $\alpha-1$ as the last letter of δ_u contributes Δ_α to $-|\gamma_u\delta_u| + 1$.

Since $\sigma_m(\beta)$ contains both $\alpha-1$ and α, there are two possible cases:

1) if $\alpha - 1$ is the last letter of $\sigma_m(\beta)$ (which is equivalent to $\beta = \alpha$), the contribution is $|\gamma'_v| - 1 + |\delta_v| + m - 1 = |\gamma'_v\delta_v| + m - 2$.

2) Otherwise, if the two letters appear consecutively, the contribution is $|\gamma'_v\delta_v| - 2$.

• **Other occurrences:** We consider two cases based on the number of the occurrences of $\alpha - 1$.

- Suppose first that $\alpha - 1$ appears exactly once in $\gamma_u\delta_u$. Consequently, α must be the first letter of γ_u, contributing $-|\gamma_u\delta_u| + 1$ to Δ_α. Thus, in this case, $\Delta_\alpha = m - |\gamma_u\delta_u| = -|\gamma_v\delta_v|$ if $\beta = \alpha$, and $\Delta_\alpha = -|\gamma_u\delta_u|$ otherwise.
- Now, assume that $\alpha - 1$ occurs for a second time in $\gamma_u\delta_u$. Since α must appear in γ_u with $\alpha - 1$, the letters must appear consecutively, with $\alpha - 1$ preceding α. These occurrences give the contribution $-|\gamma_u\delta_u| + 2$. It remains to consider the second occurrence of $\alpha - 1$ in $\gamma'_v\delta_v$. Notice that $\alpha - 1$ cannot appear in δ_v; since it cannot be the last letter of δ_v, it would be followed by a second α. Thus $\alpha - 1$ appears in γ_v. Since γ_v does not contain α, we must have that $\alpha - 1$ is the last letter of γ_v. This gives the contribution $|\gamma_v| - 1 = |\gamma'_v\delta_v| - |\delta_v| - m - 1$.

  In total, we have

$$\Delta_\alpha = |\gamma'_v\delta_v| - 2 - |\gamma_u\delta_u| + 1 - |\gamma_u\delta_u| + 2 + |\gamma'_v\delta_v| - |\delta_v| - m - 1 = -|\delta_v| - m$$

  if $\beta \neq \alpha$, and $\Delta_\alpha = -|\delta_v|$ if $\beta = \alpha$. ■

We are now ready to conclude with the proof of Lemma 7.1. The claim above asserts that the only way (4) holds is if both sides are equal to zero. In particular, this implies that $|\nu'|_\alpha = |\nu'|$, $\beta = \alpha$, and $\delta_v = \varepsilon$. Consequently, the last letter of $\sigma_m(\nu')$ is $\alpha - 1$, leading us to conclude that the words $U = \sigma_m^{k-1}(\gamma_u\, \sigma_m(u)\delta_u)$ and $V = \sigma_m^{k-1}(\gamma_v\, \sigma_m(\beta\nu'))$ both end with the last letter of $\sigma_m^{k-1}(\alpha - 1)$. This is a contradiction, in the case where $|u| \neq |v|$.

Thus, we conclude that the only possible way for $U \sim_{k+1} V$ is when $\delta_u = \delta_v = \gamma_u = \gamma_v = \varepsilon$ and $u \sim_1 v$. Hence, the proof is complete. □

8 Abelian Complexity for Short Factors

The initial values of the abelian complexity $a_{\mathbf{t}_m}(\ell)$ of $\mathbf{t}_m$, for $1 \leqslant \ell < m$ are presented in Table 4. For lengths $\ell \geqslant m$, the function is periodic with period m, i.e., $a_{\mathbf{t}_m}(\ell + m) = a_{\mathbf{t}_m}(\ell)$, and its behavior is fully described by Theorem 1.7 from [8]. Thus, the following proposition complements the findings of Chen et al.

2												
3	6											
4	10	12										
5	15	20	25									
6	21	30	39	42								
7	28	42	56	63	70							
8	36	56	76	88	100	104						
9	45	72	99	117	135	144	153					
10	55	90	125	150	175	190	205	210				
11	66	110	154	187	220	242	264	275	286			
12	78	132	186	228	270	300	330	348	366	372		
13	91	156	221	273	325	364	403	429	455	468	481	
14	105	182	259	322	385	434	483	518	553	574	595	602

Table 4: Values of $a_{\mathbf{t}_m}(\ell)$ for $1 \leqslant \ell < m \leqslant 14$; m increases from top to bottom and ℓ increases from left to right.

Proposition 8.1. *The initial values of the abelian complexity $\mathsf{a}_{\mathbf{t}_m}(\ell)$ of the generalized Thue–Morse word over m letters are given as follows.*

- *For odd $\ell < m$, say $\ell = 2\ell' + 1$, where $\ell' \geq 0$, we have*
$$\mathsf{a}_{\mathbf{t}_m}(\ell) = m\big(1 - \ell' - \ell'^2 + \ell' m\big).$$
- *For even $\ell < m$, we have*
$$\mathsf{a}_{\mathbf{t}_m}(\ell) = \frac{m}{4}\big(6 - \ell^2 - 2m + 2\ell m\big).$$

Proof. By Lemma 3.6, every pair $(i, j) \in \mathcal{A}_m^2$ appears in $\mathbf{t}_m$. Thus, any factor w of length $\ell < m$ can be written as $w = ps$, where p is a suffix of some $\sigma_m(i)$ and s is a prefix of some $\sigma_m(j)$. Our aim is to count the possible Parikh vectors for such w. Since we are dealing with abelian equivalence, and the images of a letter under σ_m are cyclic permutations of $01\cdots(m-1)$, we can limit ourselves to $|p| = \ell, \ell-1, \ldots, \lceil \ell/2 \rceil$. When p is shorter than s, we obtain exactly the same Parikh vectors.

If $|p| = \ell$, then p is of the form $t\,(t+1)\cdots(t+\ell-1)$, which is a factor of some $\sigma_m(i)$. This corresponds to the m cyclic permutations of the Parikh vector $1^\ell 0^{m-\ell}$ (expressed as a word of length m).

If $|p| = \ell - 1$ and $|s| = 1$, there are m possible suffixes of $\sigma_m(i)$ of the form $t\,(t+1)\cdots(t+\ell-2)$, where $t \in \mathcal{A}_m$. We need to determine which $j \in \mathcal{A}_m$ provides Parikh vectors that have not already been listed. Here, s is the first letter of $\sigma_m(j)$, which is j. If $j = t-1$ or $j = t+\ell-1$, then we get a Parikh vector from the first case. Thus, for j, we can choose any elements in $\mathcal{A}_m$ except these two, resulting in $m-2$ possibilities and a total of $m(m-2)$ new Parikh vectors. Note that we obtain Parikh vectors (along with their cyclic permutations) of the form $1^{\ell-1}0^r10^s$ with some isolated 1, where $r, s > 0$ and $r + s = m - \ell$, or of the form $1^r21^{\ell-r-1}0^{m-\ell}$, with one 2 in any position within the block of size $\ell - 1$.

If $|p| = \ell - 2$ and $|s| = 2$, this case is similar. We have m possible suffixes of the form $t\,(t+1)\cdots(t+\ell-3)$, where $t \in \mathcal{A}_m$. We need to determine which $j \in \mathcal{A}_m$ provides new Parikh vectors. Here, s is the first two letters of $\sigma_m(j)$, which are $j(j+1)$. If $j \in \{t-2, t-1, t+\ell-3, t+\ell-2\}$, the Parikh vectors are already described in the first two cases. Otherwise, we obtain new vectors either with a block $1^r221^{\ell-r-2}$, or with two isolated block 1^2 and $1^{\ell-2}$. This results in $m.(m-4)$ new Parikh vectors.

In general, if $|p| = \ell - u$ and $|s| = u$ with $\ell/2 > u$, then p is of the form $t\,(t+1)\cdots(t+\ell-u-1)$. To obtain new Parikh vectors, either with a block $1^s2^u1^{\ell-s-u}$, or with two isolated blocks 1^u and $1^{\ell-u}$, from $s = j(j+1)\cdots(j+u-1)$, then j cannot be in $\{t-u, \ldots, t-1, t+\ell-2u, \ldots, t+\ell-u-1\}$. Therefore, j can take $m - 2u$ values.

In conclusion, if ℓ is odd of the form $\ell = 2\ell' + 1$, we obtain a total of
$$\mathsf{a}_{\mathbf{t}_m}(\ell) = m + \sum_{u=1}^{\ell'} m(m-2u) = m\big(1 - \ell' - \ell'^2 + \ell' m\big).$$

Now, if ℓ is even, we still have to consider the situation where $|p| = |s| = \ell/2$. In this case, p and s have symmetric roles, and we should avoid double counting. We need to select two elements $i, j \in \mathcal{A}_m$ that are at distance greater than $\ell/2$ from each other (over $\mathbb{Z}/(m\mathbb{Z})$) in order to obtain Parikh vectors that are a cyclic permutation of $1^{\ell/2}0^r1^{\ell/2}0^s$, where $r, s > 0$. The number of such pairs $\{i, j\}$, where $j \notin \{i - \ell/2, \ldots, i-1, i, i+1, \ldots, i+\ell/2\}$, is given by $m(m-\ell-1)/2$. There are also m permutations of $2^{\ell/2}0^{m-\ell/2}$ when $p = s$. Hence, for even ℓ, we obtain
$$\mathsf{a}_{\mathbf{t}_m}(\ell) = m + \sum_{u=1}^{\frac{\ell}{2}-1} m(m-2u) + \frac{m(m-\ell-1)}{2} + m = \frac{m}{4}\big(6 - \ell^2 - 2m + 2\ell m\big).$$

□

Remark 8.2. Interestingly, the infinite triangular array whose initial elements are given in Table 4, exhibits several intriguing combinatorial properties and identities.

- Regarding the rows of the triangle, the following relation holds for $1 \leqslant \ell < m-4$

$$a_{\mathbf{t}_m}(\ell+4) = 2a_{\mathbf{t}_m}(\ell+3) - 2a_{\mathbf{t}_m}(\ell+1) + a_{\mathbf{t}_m}(\ell).$$

  This relation can be easily deduced from the previous proposition. For $m \geqslant 5$, the initial conditions are given by

$$(a_{\mathbf{t}_m}(1), \ldots, a_{\mathbf{t}_m}(4)) = (m, m(m+1)/2, m(m-1), m(3m-5)/2).$$

- Similarly, for each column, the following holds for all $m \geqslant 2$ and all $\ell < m$,

$$a_{\mathbf{t}_{m+3}}(\ell) = 3a_{\mathbf{t}_{m+2}}(\ell) - 3a_{\mathbf{t}_{m+1}}(\ell) + a_{\mathbf{t}_m}(\ell).$$

- Furthermore, the diagonal and parallels to the diagonal $(a_{\mathbf{t}_{\ell+2+i}}(\ell+1))_{\ell\geqslant 0}$ for all $i \geqslant 0$ satisfy the same recurrence relation of order 6

$$x_{n+6} = 2x_{n+5} + x_{n+4} - 4x_{n+3} + x_{n+2} + 2x_{n+1} - x_n.$$

- The sequence $(a_{\mathbf{t}_m}(m-2))_{m\geqslant 3} = 3, 10, 20, 39, 63, 100, 144, \ldots$ appears in several entries of the OEIS, as `A005997` (number of paraffins) and `A272764` (number of positive roots in reflection group E_n), among others.

- The sequence $(a_{\mathbf{t}_{2m+1}}(2m))_{m\geqslant 1}$ is given by $2m^3 + m^2 + 2m + 1$.

- The sequence $((a_{\mathbf{t}_{2m}}(2m-1))/2)_{m\geqslant 1} = 1, 6, 21, 52, 105, 186, 301, \ldots$ is the sequence of q-factorial numbers $([3]!_q)_{q\geqslant 0}$ where

$$[3]!_q = \frac{(1-q)(1-q^2)(1-q^3)}{(1-q)^3} = (1+q)(1+q+q^2).$$

  It appears as `A069778` in the OEIS.

9 Description of the Abelian Rauzy Graphs

The abelian Rauzy graph is defined in Definition 2.6. Refer to Section 2 for the definitions of the sets $Y_{m,L}, Y_{m,R}$, and Y_m. The aim of this section is to count the number of edges in the abelian Rauzy graph $G_{m,\ell}$ of order ℓ for $\mathbf{t}_m$, where $\ell < 2m$, as well as determine the size of the corresponding set $Y_m(\ell)$. These expressions, together with Proposition 2.7, lead to Theorem 1.13.

The structure of these graphs depends on the value of the parameter ℓ. Specifically, the behavior varies significantly depending on whether $\ell < m$ or $\ell \geqslant m$.

Example 9.1. Fig. 6 depicts the graph $G_{6,4}$. To keep clarity in the figure, we have omitted the edge labels. The color of each edge is determined by the second component of its label. Thus, two edges originating from the same vertex and sharing the same color correspond to the same element of $Y_{m,R}$. The vertices are labeled with Parikh vectors. According to Proposition 8.1, $a_{\mathbf{t}_6}(4) = 39$, which implies that the graph $G_{6,4}$ has 39 vertices. The symmetry of the graph results from Lemma 3.5.

Example 9.2. Fig. 7 depicts the graph $G_{5,4}$, which has $a_{\mathbf{t}_5}(4) = 25$ vertices. This example may help the reader follow the developments presented in the proof below, where the case of odd m and even ℓ is discussed. Providing two distinct examples is insightful. Fig. 7 exhibits a 5-fold symmetry in the graph. However, Fig. 6 shows that the three central vertices exhibit a different behavior, specifically a 3-fold symmetry, instead of the 6-fold symmetry present in the rest of the graph.

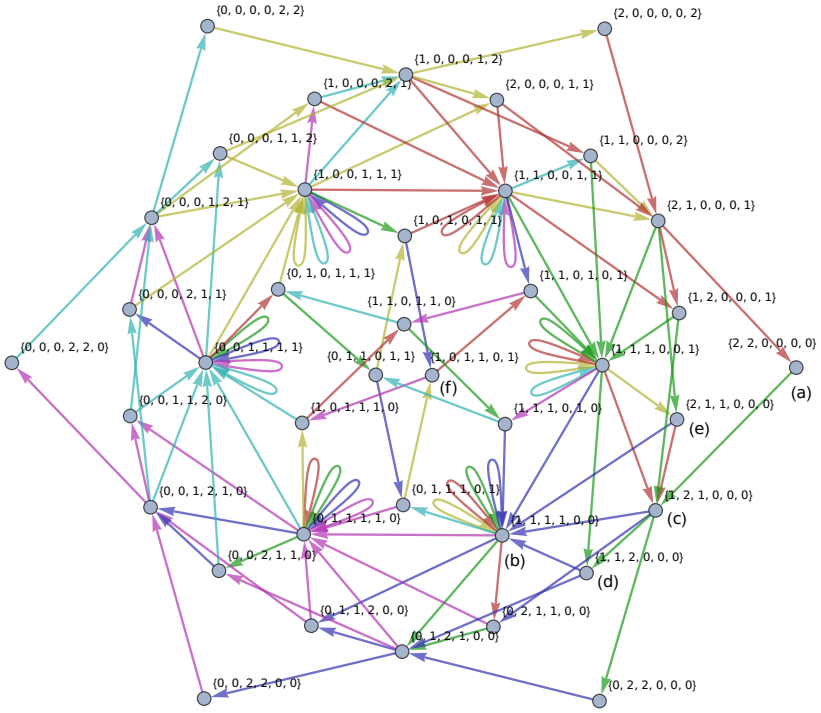


Figure 6: Abelian Rauzy graph $G_{6,4}$ of order 4 for $\mathbf{t}_6$.

9.1 When $\ell < m$

Proposition 9.3. *For $1 \leqslant \ell < m$, the number of edges in the abelian Rauzy graph $G_{m,\ell}$ is given by*

$$m(1 + \ell m - \ell).$$

Proof. For $\ell = 1$, all length-2 factors of the form ab appear in $\mathbf{t}_m$. Thus, $G_{m,1}$ is a complete directed graph with m^2 edges.

Now, assume $\ell \geqslant 2$. As a first case, let ℓ be even, in the form $2\ell'$, where $\ell' > 0$, and m is odd (as in Example 9.2). Table 5 lists the possible Parikh vectors v and their corresponding out-degree $d^+(v)$. Note that we must also consider the cyclic permutations of these vectors, which correspond to other vertices in the graph.

type	$\Psi(v)$	d^+	choices	total number of vertices of this type when ℓ even
a)	$2^{\ell'}0^{m-\ell}$	1		1
b)	$1^{\ell}0^{m-\ell}$	$\ell + m - 1$		1
c)	$1^i2^j1^{\ell-i-2j}0^{m-\ell+j}$	4	$i, j, \ell - i - 2j > 0$	$(\ell' - 1)^2$
d)	$1^{\ell-2i}2^i0^{m-\ell+i}$	2	$i, \ell - 2i > 0$	$\ell' - 1$
e)	$2^i1^{\ell-2i}0^{m-\ell+i}$	2	$i, \ell - 2i > 0$	$\ell' - 1$
f)	$1^i0^j1^{\ell-i}0^{m-\ell-j}$	2	$i, j, \ell - i, m - \ell - j > 0$	$(\ell' - \frac{1}{2})(m - \ell - 1)$

Table 5: The different types of vertices (not counting cyclic permutations).

We proceed similarly to the proof of Proposition 8.1, describing the Parikh vectors represented succinctly as words.

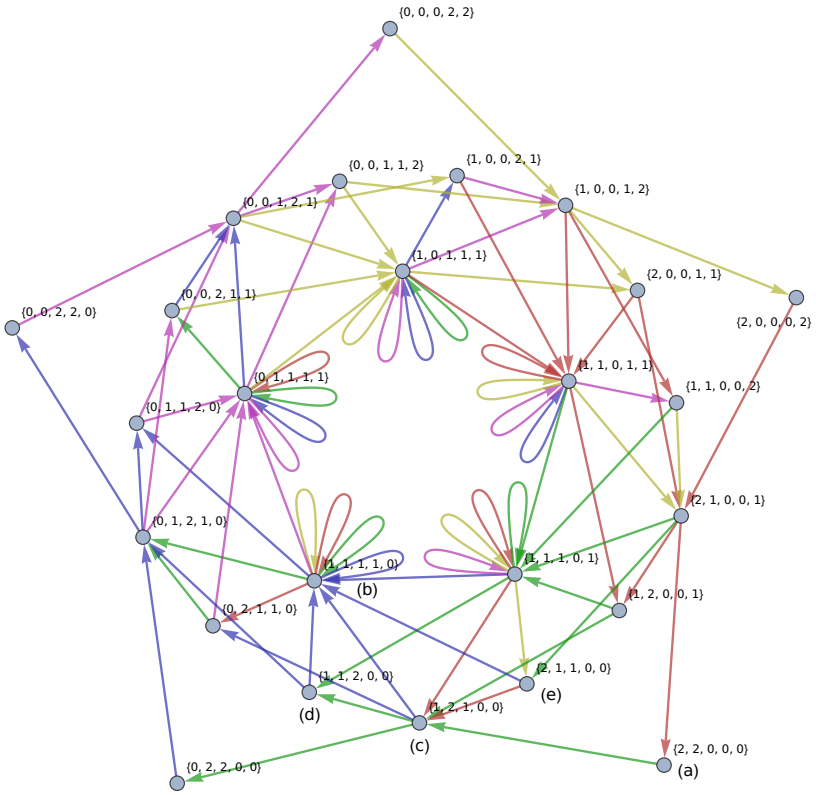


Figure 7: Abelian Rauzy graph $G_{5,4}$ of order 4 for $\mathbf{t}_5$.

(a) The factor $01\cdots\ell'01\cdots\ell'$ has a unique successor in $\mathbf{t}_m$, which is $1\cdots\ell'01\cdots\ell'(\ell'+1)$. Thus, there is an edge $2^{\ell'}0^{m-\ell} \to 12^{\ell'-1}10^{m-\ell-1}$. The reader may refer to Example 9.2 to observe the different types of vertices described in this proof. For the first type, these vertices are located on the outermost part of Fig. 7.

(b) The Parikh vector $1^{\ell}0^{m-\ell}$ can be associated with the factor $0\cdots(\ell-1)$. Since all pairs of letters occur in $\mathbf{t}_m$, the factor $0\cdots(\ell-1)a$ occurs in $\mathbf{t}_m$ for all $a \in \mathcal{A}_m$. Thus, there are m edges with the label $(0,a)$; in particular, one of them is a loop with label $(0,0)$. This Parikh vector is also associated with a factor of the form vu, where $u = 0\cdots(i-1)$ and $v = i\cdots(\ell-1)$, with $i = 1,\ldots,\ell-1$. Thus, there are $\ell-1$ loops labeled (i,i). For the second type, these vertices are located on the innermost part of Fig. 7.

(c) The Parikh vector $1^i2^j1^{\ell-i-2j}0^{m-\ell+j}$ is associated with a factor of the form uv or vu, where $u = 0\cdots(i+j-1)$ and $v = i\cdots(\ell-j-1)$. It can also be associated with a factor uv or vu, where $u = 0\cdots(\ell-j-1)$ and $v = i\cdots(i+j-1)$. This results in four edges towards the following vertices:

$$01^{i-1}2^j1^{\ell-i-2j+1}0^{m-\ell+j-1}, \quad 1^{i+1}2^j1^{\ell-i-2j-1}0^{m-\ell+j},$$
$$01^{i-1}2^{j+1}1^{\ell-i-2j-1}0^{m-\ell+j-1}, \quad 1^{i+1}2^{j-1}1^{\ell-i-2j+1}0^{m-\ell+j}.$$

(d) & (e) These cases are similar. The Parikh vector $2^i1^{\ell-2i}0^{m-\ell+i}$ is associated with a factor of the form uv or vu, where $u = 0\cdots(i-1)$ and $v = 0\cdots(\ell-i-1)$. This results in two edges labeled $(0,\ell-i)$ and $(0,i)$, which are distinct because $\ell-2i>0$.

(f) We have factors of the form uv or vu, where $u = 0\cdots(i-1)$ and $v = (i+j)\cdots(\ell+j-1)$. This results in two edges labeled $(0,\ell+j)$ and $(i+j,i)$.

Next, we count the total number of edges. To do so, we need to determine the number of vertices of each type. There are m pairwise distinct cyclic permutations of the vector of type (a). The same observation applies for type (b). This results in $m + m(\ell + m - 1) = m(\ell + m)$ edges in $G_{m,\ell}$.

For a vector of type (c), for each valid $j \leqslant \ell' - 1$, there are $\ell - 2j - 1$ ways to arrange $\ell - 2j$ ones on both sides of 2^j. This results in

$$\sum_{j=1}^{\ell'-1} (\ell - 2j - 1) = (\ell' - 1)^2. \tag{5}$$

Taking into account the cyclic permutations, we obtain $4m(\ell' - 1)^2$ edges.

For a vector of type (d) or (e), there are $\ell' - 1$ choices for i. This resulting in a total of $4m(\ell' - 1)$ edges.

The type (f) requires extra caution: since ℓ is even, not all cyclic permutations are distinct, so we must avoid double counting. We have to limit ourselves to $i \leqslant \ell'$. Indeed, the m cyclic permutations of $1^i 0^j 1^{\ell-i} 0^{m-\ell-j}$ and those of $1^{\ell-i} 0^{m-\ell-j} 1^i 0^j$ are identical. For each $i < \ell'$, there are $m - \ell - 1$ choices for j. This results in $2m(\ell' - 1)(m - \ell - 1)$ edges. When $i = \ell'$, there are two blocks of ones of the same size, giving only $(m - \ell - 1)/2$ choices for j. This is the only place where the fact that m is odd plays a role. This provides $2m(m - \ell - 1)/2 = m(m - \ell - 1)$ edges. Summing up all contributions yields the expected value

$$\begin{aligned} m(\ell + m) + 4m(\ell' - 1)^2 + 4m(\ell' - 1) + 2m(\ell' - 1)(m - \ell - 1) \\ + m(m - \ell - 1) = m(1 + \ell m - \ell). \end{aligned}$$

If m is even and $i = \ell'$, we must consider $j < (m - \ell)/2$ and $j = (m - \ell)/2$ separately because in the latter case, there are also two blocks of zeroes of the same size. Thus, we must again avoid double counting. This results in $2m((m - \ell/2) - 1) + m/2$ edges. The last term corresponds to the permutations of $1^{\ell'} 0^{(m-\ell)/2} 1^{\ell'} 0^{(m-\ell)/2}$, which can be observed in Fig. 6 with the three innermost vertices. The summation yields the same expression.

The case where ℓ is odd treated similarly. Note that there are no Parikh vectors of type (a). □

Remark 9.4. For $1 \leqslant \ell < m$, the graph $G_{m,\ell}$ is an Eulerian graph. The previous proof can be reproduced by focusing on the in-degree of the vertices and show that for all vertices v, $d^+(v) = d^-(v)$. Since $\mathbf{t}_m$ is recurrent, the graph $G_{m,\ell}$ is strongly connected. This suffices to conclude.

Proposition 9.5. *For $1 \leqslant \ell < m$, the following holds*

$$\#Y_{m,R}(\ell) = \#Y_{m,L}(\ell) = m(1 + \ell m - \ell) - \frac{m}{2}\ell(\ell - 1).$$

In particular, the value of $\#Y_m(\ell)$ is given by

$$\#Y_m(\ell) = 2m(1 + \ell m - \ell) - m\ell(\ell - 1).$$

Proof. Assume ℓ is even, of the form $2\ell'$. To compute $\#Y_{m,R}(\ell)$, we must identify the edges in $G_{m,\ell}$ that are outgoing from a vertex with labels sharing the same second component. If such edges exist, they are counted once in $\#Y_{m,R}(\ell)$. Our strategy is to subtract, from the total number of edges given by Proposition 9.3, those that do not contribute a new element to the set $\#Y_{m,R}(\ell)$. In Example 9.1, to compute $\#Y_{6,R}(4)$, one must sum, for each vertex, the number of outgoing edges, counting only one edge per distinct color.

Using the same notation as in the proof of Proposition 9.3, only vertices of type (b), (c), or (d) will contribute. We now identify the edges whose labels share the same second component. The vertex $1^\ell 0^{m-\ell}$ has $\ell - 1$ outgoing edges labeled $(0, j)$ and $\ell - 1$ loops labeled (j, j), for $j = 1, \ldots, \ell - 1$. (Refer to Figs. 6 and 7 to observe the vertices having loops.) Considering the cyclic permutations

of the Parikh vector, we must subtract $m(\ell-1)$ from the total number of edges. A vertex of type (c) has 2 has two outgoing edges with second component $\ell-j$, and two outgoing edges with second component $i+j$. (Refer to Figs. 6 and 7 to observe the vertices with an out-degree of 4.) Moreover, $\ell-j \neq i+j$ since $\ell-i-2j>0$. From (5), we must subtract $2m(\ell'-1)^2$. Finally a vertex of type (d) has two outgoing edges with second component $\ell-i$. Hence, we subtract $m(\ell'-1)$. The total amount to subtract is:

$$m\big[\ell-1+2(\ell'-1)^2+\ell'-1\big] = \frac{m\ell(\ell-1)}{2}.$$

The remaining cases are treated similarly.

To determine $\#Y_{m,L}(\ell)$, we need to identify the edges in $G_{m,\ell}$ that are incoming to a vertex with labels sharing the same first component. If such edges exist, they are counted once in $Y_{m,L}(\ell)$. Only vertices of type (b), (c), or (e) contribute. Refer to Example 9.6 for further clarification. The reasoning is similar in this case. □

Example 9.6. Fig. 8 depicts the graph $G_{5,4}$. Compared to Examples 9.1 and 9.2, the color of each edge is determined by the first component of its label. Vertices are labeled with their corresponding Parikh vectors.

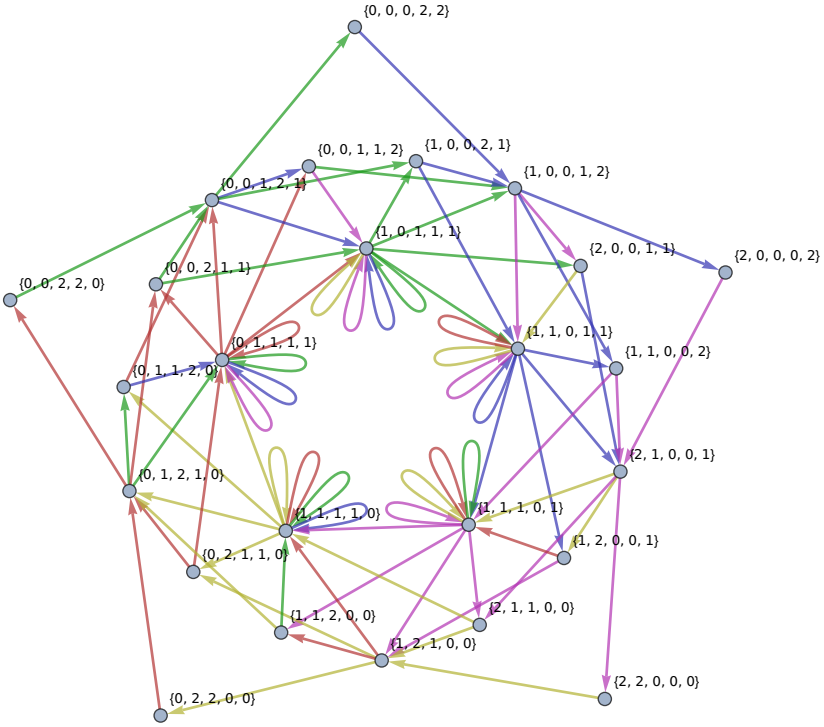


Figure 8: Abelian Rauzy graph $G_{5,4}$ of order 4 for $\mathbf{t}_5$; edges colored by the first component of the label.

9.2 When $\ell \geqslant m$

Proposition 9.7. *For $m \leqslant \ell < 2m$, the number of edges in the abelian Rauzy graph $G_{m,\ell}$ is given by*

$$m(m^2-m+1).$$

Proof. Let $b \in \mathcal{A}_m$. Due to the symmetry of σ_m, we count the number of edges labeled $(0,b)$ and then multiply the result by m. So, we focus on factors of length $\ell+1$ that start with 0 and end with b. These factors can be of one of the following two forms

- $u v b$, where u starts with 0, $|u| = t \leqslant m$, and $|v| = \ell - t < m$, i.e., $\ell - m < t \leqslant m$; or
- $u\, \sigma_m(a) v b$, for some letter a, and where u starts with 0, $|u| = t \leqslant \ell - m$, and $|v| = \ell - m - t$, i.e., $1 \leqslant t \leqslant \ell - m$.

In both cases, u (respectively, vb) is a suffix (respectively, prefix) of the image of a letter under σ_m. In particular, all letters of u are determined by the first letter 0, and all letters of v are determined by b. Note that the first letter of v is congruent to $b - \ell + t$ modulo m.

Consider the first case, where $\ell - b = m$. There is a single edge labeled $(0, b)$ from $2^b 1^{m-b}$ to $12^b 1^{m-b-1}$. Since $|u| = t$, the last letter of u is $t - 1$. Under the assumption $\ell - b = m$, the first letter of v is t. Therefore, all the previously described factors have the same Parikh vector.

Next, assume that $\ell - b \neq m$. We will prove that there are m pairwise distinct Parikh vectors, each with an outgoing edge labeled $(0, b)$. Since there are $m - 1$ possible values for b, we obtain the expected value of $m(m-1) = m^2 - m$. In this case, the last letter of u is $t - 1$, and the first letter of v is $b - \ell - t$ which is not congruent to t modulo m.

First, assume that we have two factors uvb and $u'v'b$ of the first form, where $|u'| = t' < |u| = t \leqslant m$. Then, $\Psi(uvb) - \Psi(u'v'b)$, and also contains 1's in positions corresponding to $t', t'+1, \ldots, t-1$ and contains -1's in positions corresponding to $b - \ell + t', \ldots, b - \ell + t - 1$ (modulo m). Since $\ell - b \neq m$, the two intervals of length $t - t'$, made of these positions are not equal over $\mathbb{Z}/(m\mathbb{Z})$. Therefore, $\Psi(uvb) - \Psi(u'v'b) \neq 0$.

A similar reasoning applies to the two factors $u\, \sigma_m(a) v b$ and $u'\, \sigma_m(a') v' b$ of the second form.

Finally, we compare a factor $x = uvb$ of the first form with a factor $y = u'\, \sigma_m(a) v' b$ of the second form. Let $t = |u|$ and $t' = |u'|$, with $\ell - m < t \leqslant m$ and $0 < t' \leqslant \ell - m$. Then, x and y have the same prefix (respectively, suffix) of length t' (respectively, $\ell - m - t'$). Thus,

$$\Psi(x) - \Psi(y) = \Psi(t'(t'+1)\cdots(t-1)(b-\ell+t)\cdots(b-t')) - \Psi(\sigma_m(a)).$$

This difference is non-zero, as $\ell - b \neq m$. Consequently, the length-m word

$$t'(t'+1)\cdots(t-1)(b-\ell+t)\cdots(b-t')$$

contains at least one repeated letter. □

Example 9.8. In Fig. 9, we have depicted the graph $G_{4,5}$. The color of each edge is determined by the first component of its label, as the next proof focuses on the set $Y_{m,L}$. The vertices are labeled with their corresponding Parikh vectors.

Proposition 9.9. *For $m \leqslant \ell < 2m$, the following holds*

$$\#Y_{m,R} = \#Y_{m,L} = \frac{m + m^2(m-1)}{2}.$$

In particular, $\#Y_m(\ell)$ is given by

$$\#Y_m(\ell) = 2m + m^2(m-1).$$

Proof. We focus on $Y_{m,L}$, using the same notation as in the proof of Proposition 9.7. The strategy is similar to that used in Proposition 9.5: subtracting, from the total number of edges given by Proposition 9.7, those that do not contribute a new element to the set $\#Y_{m,L}(\ell)$.

If $\ell - b = m$, there are m incoming edges labeled as $(0, i)$ for all $i \in \mathcal{A}_m$, directed to $x = 12^b 1^{m-b-1}$. The initial vertices $\Psi(0) + x - \Psi(i)$ are pairwise distinct. So we have to subtract $m - 1$ from the total number of edges in $G_{m,\ell}$. For example, observe the four yellow vertices leading to vertex 1211 in Fig. 9. For distinct $b, b' \neq \ell - m$, there exists a unique Parikh vector $x_{\{b,b'\}}$ with two incoming edges labeled as $(0, b)$ and $(0, b')$. For two such pairs $\{b, b'\}$ and $\{c, c'\}$, the corresponding vertices are such that $x_{\{b,b'\}} \neq x_{\{c,c'\}}$. Note that the number of these pairs is $\binom{m-1}{2}$. In Fig. 9, three vertices — namely, 0221, 1121 and 1112, each have two yellow incoming edges. So we also have to subtract $(m-1)(m-2)/2$. Thus,

$$\#Y_{m,L} = m(m^2 - m + 1) - m\left[m - 1 + \frac{(m-1)(m-2)}{2}\right] = m\left(\frac{(m^2-m)}{2} + 1\right).$$

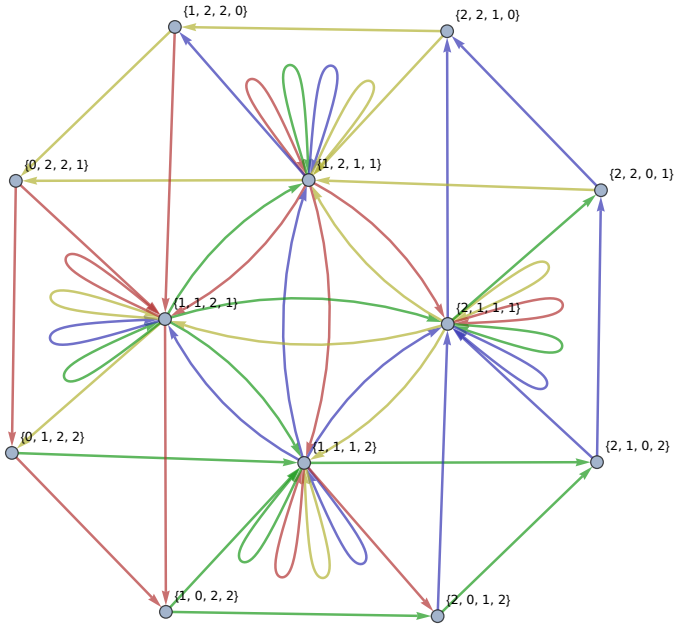


Figure 9: Abelian Rauzy graph $G_{4,5}$ of order 5 for $\mathbf{t}_4$; edges colored by the first component of the label.

To obtain the result for $Y_{m,R}$, the reasoning remains identical; however, one has to consider edges labeled as $(b, 0)$. □

Remark 9.10. For all $j \geqslant 1$ and $m \leqslant \ell < 2m$, the abelian Rauzy graph $G_{m,\ell+j\cdot m}$ is isomorphic to $G_{m,\ell}$. Refer to the proof of Proposition 9.7. We may have factors of length ℓ in one of the following two forms

- uv where u starts with $|u| = t \leqslant m$, $|v| = \ell - t < m$, i.e., $\ell - m < t \leqslant m$; or
- $u'\,\sigma_m(c)v'$ for some letter c, where $|u'| = t \leqslant \ell - m$ and $|v'| = \ell - m - t$, i.e., $1 \leqslant t \leqslant \ell - m$.

In both cases, u, u' (respectively, v, v') is a suffix of $\sigma_m(a)$ for some letter a (respectively, prefix of $\sigma_m(b)$ for some letter b). By Lemma 3.6, there exists a factor x (respectively, y) of length j (respectively, $j+1$) such that $u\,\sigma_m(x)v$ and $u'\,\sigma_m(y)v'$ are factors of $\mathbf{t}_m$. Note that

$$\Psi(u\,\sigma_m(x)v) = \Psi(uv) + j\cdot(1,\ldots,1)$$

and

$$\Psi(u'\,\sigma_m(y)v') = \Psi(u'\,\sigma_m(c)v') + j\cdot(1,\ldots,1).$$

These two observations show that $G_{m,\ell+t\cdot m}$ and $G_{m,\ell}$ are the same graph up to a renaming of the vertices.

The careful reader may observe that this remark provides an alternative proof of our main result, Theorem 1.14. Once the structure of the abelian Rauzy graphs is well understood, the formula given by Proposition 2.7 also provides a characterization of the k-binomial complexity. The two approaches developed in this article are, in our view, complementary. Each approach provides its own set of combinatorial perspectives. With this article, we have reconciled several approaches. First, we simplified Lejeune's arguments in [16] and considered the same type of equivalence relation for larger alphabets. Next, we applied abelian Rauzy graphs in a different context from that in [29].

10 Proof of Lemma 4.4

Recall from Section 4 that $\overline{a}$ denotes $-a$ for $a \in \mathbb{Z}$. Lemma 4.4 is crucial for proving Lemma 7.1.

Proof. Let $e = 0\overline{1}\cdots\overline{k}$. The subword e may appear entirely in $\sigma_m^k(u)$, entirely in $\sigma_m^{k-1}(\gamma\delta)$, or intersects both parts. So we have

$$\binom{\sigma_m^{k-1}(\gamma\,\sigma_m(u)\delta)}{e} = \binom{\sigma_m^k(u)}{e} + \binom{\sigma_m^{k-1}(\gamma\delta)}{e} + \sum_{\substack{e=xyz\\ 0<|y|<k+1}} \binom{\sigma_m^{k-1}(\gamma)}{x}\binom{\sigma_m^k(u)}{y}\binom{\sigma_m^{k-1}(\delta)}{z}.$$

Since $|u| = |u'|$, by Proposition 3.7, $\sigma_m^k(u) \sim_k \sigma_m^k(u')$ and

$$\begin{aligned} &\binom{\sigma_m^{k-1}(\gamma\,\sigma_m(u)\delta)}{e} - \binom{\sigma_m^{k-1}(\gamma'\,\sigma_m(u')\delta')}{e} && (6)\\ = \;&\binom{\sigma_m^k(u)}{e} - \binom{\sigma_m^k(u')}{e} + \binom{\sigma_m^{k-1}(\gamma\delta)}{e} - \binom{\sigma_m^{k-1}(\gamma'\delta')}{e} \\ &+ \sum_{\substack{e=xyz\\ 0<|y|<k+1}} \binom{\sigma_m^k(u)}{y}\left[\binom{\sigma_m^{k-1}(\gamma)}{x}\binom{\sigma_m^{k-1}(\delta)}{z} - \binom{\sigma_m^{k-1}(\gamma')}{x}\binom{\sigma_m^{k-1}(\delta')}{z}\right]. && (7)\end{aligned}$$

Observing that the factors x, y, and z in the above sum are respectively of the form $x = 0\overline{1}\cdots\overline{j-1}$; $y = \overline{j}\cdots\overline{j+\ell-1}$; $z = \overline{j+\ell}\cdots\overline{k}$ for $1 \leqslant \ell \leqslant k$, let us rewrite term (7) of the latter expression as

$$\sum_{\ell=1}^{k}\sum_{j=0}^{k-\ell+1} \binom{\sigma_m^k(u)}{\overline{j}\cdots\overline{j+\ell-1}}\left[\binom{\sigma_m^{k-1}(\gamma)}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta)}{\overline{j+\ell}\cdots\overline{k}} - \binom{\sigma_m^{k-1}(\gamma')}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta')}{\overline{j+\ell}\cdots\overline{k}}\right].$$

By Lemma 4.3, the coefficient $\binom{\sigma_m^k(u)}{\overline{j}\cdots\overline{j+\ell-1}}$ equals $\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{\ell-1}}$ for each j since $\ell \leqslant k$; thus, the sum simplifies to

$$\sum_{\ell=1}^{k}\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{\ell-1}}\sum_{j=0}^{k-\ell+1}\left[\binom{\sigma_m^{k-1}(\gamma)}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta)}{\overline{j+\ell}\cdots\overline{k}} - \binom{\sigma_m^{k-1}(\gamma')}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta')}{\overline{j+\ell}\cdots\overline{k}}\right].$$

By Lemma 4.3 again, we may replace $\binom{\sigma_m^{k-1}(\delta)}{\overline{j+\ell}\cdots\overline{k}}$ with $\binom{\sigma_m^{k-1}(\delta)}{\overline{j}\cdots\overline{k-\ell}}$ and $\binom{\sigma_m^{k-1}(\delta')}{\overline{j+\ell}\cdots\overline{k}}$ with $\binom{\sigma_m^{k-1}(\delta')}{\overline{j}\cdots\overline{k-\ell}}$, as long as $|\overline{j}\cdots\overline{k-\ell}| < k$, i.e., when $\ell \geqslant 2$ or $\ell = 1$ and $j \geqslant 1$. We decompose the sum accordingly (for convenience, we also add and subtract the same extra term)

$$\begin{aligned}&\sum_{\ell=2}^{k}\binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{\ell-1}}\sum_{j=0}^{k-\ell+1}\left[\binom{\sigma_m^{k-1}(\gamma)}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta)}{\overline{j}\cdots\overline{k-\ell}} - \binom{\sigma_m^{k-1}(\gamma')}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta')}{\overline{j}\cdots\overline{k-\ell}}\right]\\ &+\binom{\sigma_m^k(u)}{0}\left(\sum_{j=1}^{k}\left[\binom{\sigma_m^{k-1}(\gamma)}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta)}{\overline{j}\cdots\overline{k-1}} - \binom{\sigma_m^{k-1}(\gamma')}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(\delta')}{\overline{j}\cdots\overline{k-1}}\right]\right.\\ &+\binom{\sigma_m^{k-1}(\delta)}{0\overline{1}\cdots\overline{k-1}} - \binom{\sigma_m^{k-1}(\delta')}{0\overline{1}\cdots\overline{k-1}}\\ &\left.+\binom{\sigma_m^{k-1}(\delta)}{\overline{1}\cdots\overline{k}} - \binom{\sigma_m^{k-1}(\delta')}{\overline{1}\cdots\overline{k}} - \left[\binom{\sigma_m^{k-1}(\delta)}{0\overline{1}\cdots\overline{k-1}} - \binom{\sigma_m^{k-1}(\delta')}{0\overline{1}\cdots\overline{k-1}}\right]\right).\end{aligned}$$

Since

$$\sum_{j=0}^{k-\ell+1}\binom{\sigma_m^{k-1}(x)}{0\overline{1}\cdots\overline{j-1}}\binom{\sigma_m^{k-1}(y)}{\overline{j}\cdots\overline{k-\ell}} = \binom{\sigma_m^{k-1}(xy)}{0\overline{1}\cdots\overline{k-\ell}}$$

for any words x, y, we further simplify to

$$\sum_{\ell=1}^{k} \binom{\sigma_m^k(u)}{0\overline{1}\cdots\overline{\ell-1}} \left[\binom{\sigma_m^{k-1}(\gamma\delta)}{0\overline{1}\cdots\overline{k-\ell}} - \binom{\sigma_m^{k-1}(\gamma'\delta')}{0\overline{1}\cdots\overline{k-\ell}} \right] + m^{k-1}|u| \left(\binom{\sigma_m^{k-1}(\delta)}{\overline{1}\cdots\overline{k}} - \binom{\sigma_m^{k-1}(\delta)}{0\cdots\overline{k-1}} - \binom{\sigma_m^{k-1}(\delta')}{\overline{1}\cdots\overline{k}} + \binom{\sigma_m^{k-1}(\delta')}{0\cdots\overline{k-1}} \right). \quad (8)$$

Now $\binom{\sigma_m^{k-1}(\delta)}{\overline{1}\cdots\overline{k}} = \binom{\sigma_m^{k-1}\tau_m(\delta)}{0\cdots\overline{k-1}}$, where we recall that τ_m is the morphism defined by $\tau_m(i) = i+1$. Thus, by Corollary 4.2, the second term in (8) simplifies to:

$$\begin{aligned} m^{k-1}|u| \Big(m^{\binom{k-1}{2}} (|\delta+1|_0 - |\delta|_0 - |\delta'+1|_0 + |\delta'|_0) \Big) \\ = m^{\binom{k}{2}} |u| \big(|\delta|_{\overline{1}} - |\delta|_0 - |\delta'|_{\overline{1}} + |\delta'|_0 \big). \end{aligned} \quad (9)$$

Consider the sum appearing in (8). Since $|\delta\gamma| = |\delta'\gamma'|$, by Proposition 3.7, $\sigma_m^{k-1}(\gamma\delta) \sim_{k-1} \sigma_m^{k-1}(\gamma'\delta')$, and the sum reduces to a single term (corresponding to $\ell = 1$)

$$\binom{\sigma_m^k(u)}{0} \left[\binom{\sigma_m^{k-1}(\gamma\delta)}{0\overline{1}\cdots\overline{k-1}} - \binom{\sigma_m^{k-1}(\gamma'\delta')}{0\overline{1}\cdots\overline{k-1}} \right] = m^{k-1}|u| \, (|\gamma\delta|_0 - |\gamma'\delta'|_0) m^{\binom{k-1}{2}}$$

(where we have used Corollary 4.2) and is equal to

$$|u| \, m^{\binom{k}{2}} (|\gamma\delta|_0 - |\gamma'\delta'|_0).$$

We can now return to the initial difference (6) of interest. By applying Corollary 4.2 again, we get that (6) is equal to

$$\begin{aligned} &\binom{\sigma_m^{k-1}(\gamma\delta)}{e} - \binom{\sigma_m^{k-1}(\gamma'\delta')}{e} + \\ &m^{\binom{k}{2}} \Big[|u|_0 - |u'|_0 + |u| \left(|\gamma\delta|_0 - |\gamma'\delta'|_0 + |\delta|_{\overline{1}} - |\delta|_0 - |\delta'|_{\overline{1}} + |\delta'|_0 \right) \Big]. \end{aligned}$$

To conclude the proof, we develop the difference between the first two terms. Let $\gamma\delta = x_1 \cdots x_t$ and $\gamma'\delta' = x'_1 \cdots x'_t$. We use the same argument as in the proof of Proposition 4.1. We need to count occurrences of the subword e. If an occurrence is split across multiple m^{k-1}-blocks and at most $k-1$ letters appear in any block, then these occurrences will cancel because $\sigma_m^{k-1}(x_i) \sim_{k-1} \sigma_m^{k-1}(x'_i)$. We only have to consider occurrences where at least k letters (out of $k+1$) appear in the same m^{k-1}-block. Then, we look at e occurring entirely within one m^{k-1}-block, given by the following expression

$$\sum_{i=1}^{t} \left(\binom{\sigma_m^{k-1}(x_i)}{e} - \binom{\sigma_m^{k-1}(x'_i)}{e} \right)$$

and this sum vanishes because $\gamma\delta \sim_1 \gamma'\delta'$. Alternatively, e is split with k letters in one m^{k-1}-block and one (the first or the last) in another m^{k-1}-block, we obtain

$$\begin{aligned} &\sum_{i=1}^{t-1} \sum_{j=i+1}^{t} \left(\binom{\sigma_m^{k-1}(x_i)}{0} \binom{\sigma_m^{k-1}(x_j)}{\overline{1}\cdots\overline{k}} - \binom{\sigma_m^{k-1}(x'_i)}{0} \binom{\sigma_m^{k-1}(x'_j)}{\overline{1}\cdots\overline{k}} \right) \\ + &\sum_{i=1}^{t-1} \sum_{j=i+1}^{t} \left(\binom{\sigma_m^{k-1}(x_i)}{0\,\overline{1}\cdots\overline{k-1}} \binom{\sigma_m^{k-1}(x_j)}{\overline{k}} - \binom{\sigma_m^{k-1}(x'_i)}{0\,\overline{1}\cdots\overline{k-1}} \binom{\sigma_m^{k-1}(x'_j)}{\overline{k}} \right). \end{aligned}$$

We get

$$\sum_{i=1}^{t-1}\sum_{j=i+1}^{t} m^{k-2}\left(\binom{\sigma_m^{k-1}(x_j+1)}{0\overline{1}\cdots\overline{k-1}}-\binom{\sigma_m^{k-1}(x'_j+1)}{0\overline{1}\cdots\overline{k-1}}\right)$$
$$+\sum_{i=1}^{t-1}\sum_{j=i+1}^{t} m^{k-2}\left(\binom{\sigma_m^{k-1}(x_i)}{0\,\overline{1}\cdots\overline{k-1}}-\binom{\sigma_m^{k-1}(x'_i)}{0\,\overline{1}\cdots\overline{k-1}}\right).$$

By Corollary 4.2, it is equal to

$$\sum_{i=1}^{t-1}\sum_{j=i+1}^{t} m^{k-2}m^{\binom{k-1}{2}}(|x_j|_{\overline{1}}-|x'_j|_{\overline{1}})+\sum_{i=1}^{t-1}\sum_{j=i+1}^{t} m^{k-2}m^{\binom{k-1}{2}}(|x_i|_0-|x'_i|_0)$$

which can be rewritten as

$$m^{k-2}m^{\binom{k-1}{2}}\sum_{j=2}^{t}(j-1)\left(|x_j|_{\overline{1}}-|x'_j|_{\overline{1}}\right)+m^{k-2}m^{\binom{k-1}{2}}\sum_{i=1}^{t-1}(t-i)(|x_i|_0-|x'_i|_0).$$

If $x_j=\overline{1}$, the factor $j-1$ represents the number of letters to the left of x_j and if $x_i=0$, the factor $t-i$ represents the number of letters to the right of x_i. Therefore, we can write

$$m^{k-2}m^{\binom{k-1}{2}}\sum_{b\in\mathcal{A}_m}\left(\binom{\gamma\delta}{b\overline{1}}-\binom{\gamma'\delta'}{b\overline{1}}+\binom{\gamma\delta}{0b}-\binom{\gamma'\delta'}{0b}\right).$$

□

Acknowledgments

We warmly thank the two anonymous referees for their valuable advices improving the presentation of this paper.

References

[1] Boris Adamczewski. Balances for fixed points of primitive substitutions. *Theoret. Comput. Sci.*, 307(1):47–75, 2003. Words. doi:10.1016/S0304-3975(03)00092-6.

[2] Jean-Paul Allouche. Thue, Combinatorics on words, and conjectures inspired by the Thue-Morse sequence. *Journal de théorie des nombres de Bordeaux*, 27(2):375–388, 2015. URL: http://www.numdam.org/articles/10.5802/jtnb.906/, doi:10.5802/jtnb.906.

[3] Jean-Paul Allouche and Jeffrey Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In *Sequences and their applications (Singapore, 1998)*, Springer Ser. Discrete Math. Theor. Comput. Sci., pages 1–16. Springer, London, 1999.

[4] Jean-Paul Allouche and Jeffrey Shallit. Sums of digits, overlaps, and palindromes. *Discrete Mathematics & Theoretical Computer Science*, 4, 2000. doi:10.46298/dmtcs.282.

[5] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences: Theory, applications, generalizations*. Cambridge University Press, Cambridge, 2003. doi:10.1017/CBO9780511546563.

[6] Mélodie Andrieu and Léo Vivion. Personal communication.

[7] Ľubomíra Balková. Factor frequencies in generalized Thue-Morse words. *Kybernetika*, 48(3):371–385, 2012.

[8] Jin Chen and Zhi-Xiong Wen. On the abelian complexity of generalized Thue-Morse sequences. *Theor. Comput. Sci.*, 780:66–73, 2019. doi:10.1016/j.tcs.2019.02.014.

[9] MichaĹ Dębski, Jarosł aw Grytczuk, Barbara Nayar, Urszula Pastwa, Joanna Sokół, Michał Tuczyński, Przemysł aw Wenus, and Krzysztof Węsek. Avoiding multiple repetitions in Euclidean spaces. *SIAM J. Discrete Math.*, 34(1):40–52, 2020. doi:10.1137/18M1180347.

[10] Jean-Marie Dumont and Alain Thomas. Systèmes de numération et fonctions fractales relatifs aux substitutions. (Numeration systems and fractal functions related to substitutions). *Theor. Comput. Sci.*, 65(2):153–169, 1989. doi:10.1016/0304-3975(89)90041-8.

[11] Anna E. Frid. On the frequency of factors in a D0L word. *J. Autom. Lang. Comb.*, 3(1):29–41, 1998.

[12] Ying-Jun Guo, Xiao-Tao Lü, and Zhi-Xiong Wen. On the boundary sequence of an automatic sequence. *Discrete Math.*, 345(1):9, 2022. Id/No 112632. doi:10.1016/j.disc.2021.112632.

[13] L. Kennard, M. Zaremsky, and J. Holdener. Generalized Thue-Morse sequences and the von Koch curve. *Int. J. Pure Appl. Math.*, 47(3):397–403, 2008.

[14] M. Kolář, M. K. Ali, and Franco Nori. Generalized thue-morse chains and their physical properties. *Phys. Rev. B*, 43:1034–1047, Jan 1991. doi:10.1103/PhysRevB.43.1034.

[15] Jakub Kozik and Piotr Micek. Nonrepetitive choice number of trees. *SIAM J. Discrete Math.*, 27(1):436–446, 2013. doi:10.1137/120866361.

[16] Marie Lejeune, Julien Leroy, and Michel Rigo. Computing the k-binomial complexity of the Thue-Morse word. *J. Comb. Theory, Ser. A*, 176:44, 2020. Id/No 105284. doi:10.1016/j.jcta.2020.105284.

[17] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997. With a foreword by Roger Lyndon and a preface by Dominique Perrin, Corrected reprint of the 1983 original, with a new preface by Perrin. doi:10.1017/CBO9780511566097.

[18] Xiao-Tao Lü, Jin Chen, Zhi-Xiong Wen, and Wen Wu. On the 2-binomial complexity of the generalized Thue-Morse words. *Theor. Comput. Sci.*, 986:14, 2024. Id/No 114342. doi:10.1016/j.tcs.2023.114342.

[19] László Mérai and Arne Winterhof. On the pseudorandomness of automatic sequences. *Cryptogr. Commun.*, 10(6):1013–1022, 2018. doi:10.1007/s12095-017-0260-7.

[20] Harold Marston Morse. Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.*, 22(1):84–100, 1921. doi:10.2307/1988844.

[21] Ignacio Palacios-Huerta. Tournaments, fairness and the Prouhet-Thue-Morse sequence. *Economic inquiry*, 50:848–849, 2012.

[22] Olga G. Parshina. On arithmetic index in the generalized Thue-Morse word. In *Combinatorics on words*, volume 10432 of *Lecture Notes in Comput. Sci.*, pages 121–131. Springer, Cham, 2017. doi:10.1007/978-3-319-66396-8_12.

[23] Michel Rigo. *Formal languages, automata and numeration systems. Vol. 2. Applications to recognizability and decidability*. Hoboken, NJ: John Wiley & Sons; London: ISTE, 2014. doi:10.1002/9781119042853.

[24] Michel Rigo. Relations on words. *Indag. Math. (N.S.)*, 28(1):183–204, 2017. doi:10.1016/j.indag.2016.11.018.

[25] Michel Rigo and P. Salimov. Another generalization of abelian equivalence: binomial complexity of infinite words. *Theor. Comput. Sci.*, 601:47–57, 2015. doi:10.1016/j.tcs.2015.07.025.

[26] Michel Rigo, Manon Stipulanti, and Markus Whiteland. Biomial coefficients of words and beyond. To appear in Combinatorics, Dynamics, and Formal Methods.

[27] Michel Rigo, Manon Stipulanti, and Markus A. Whiteland. Automaticity and Parikh-collinear morphisms. In Anna Frid and Robert Mercaş, editors, *Combinatorics on Words*, pages 247–260, Cham, 2023. Springer Nature Switzerland. doi:10.1007/978-3-031-33180-0_19.

[28] Michel Rigo, Manon Stipulanti, and Markus A. Whiteland. Automatic abelian complexities of Parikh-collinear fixed points. *Theory Comput Syst*, 2024. doi:10.1007/s00224-024-10197-5.

[29] Michel Rigo, Manon Stipulanti, and Markus A. Whiteland. Characterizations of families of morphisms and words via binomial complexities. *Eur. J. Comb.*, 118:35, 2024. Id/No 103932. doi:10.1016/j.ejc.2024.103932.

[30] S. Sahel, R. Amri, D. Gamra, M. Lejeune, M. Benlahsen, K. Zellama, and H. Bouchriha. Effect of sequence built on photonic band gap properties of one-dimensional quasi-periodic photonic crystals: Application to thue-morse and double-period structures. *Superlattices and Microstructures*, 111:1–9, 2017. URL: https://doi.org/10.1016/j.spmi.2017.04.031.

[31] Patrice Séébold. On some generalizations of the Thue-Morse morphism. *Theoret. Comput. Sci.*, 292(1):283–298, 2003. Selected papers in honor of Jean Berstel. doi:10.1016/S0304-3975(01)00228-6.

[32] Štěpán Starosta. Generalized Thue-Morse words and palindromic richness. *Kybernetika (Prague)*, 48(3):361–370, 2012.

[33] Léo Vivion. New examples of words for which the binomial complexities and the subword complexity coincide, 2025. arXiv:2509.11172.

[34] Janusz Wolny, Anna Wnęk, and Jean-Louis Verger-Gaugry. Fractal behaviour of diffraction pattern of thue–morse sequence. *Journal of Computational Physics*, 163(2):313–327, 2000. URL: https://doi.org/10.1006/jcph.2000.6563.

[35] E. M. Wright. Prouhet's 1851 solution of the Tarry-Escott problem of 1910. *Amer. Math. Monthly*, 66:199–201, 1959. doi:10.2307/2309513.