COUNTING THE NUMBER OF NON-ZERO COEFFICIENTS IN ROWS OF GENERALIZED PASCAL TRIANGLES

JULIEN LEROY, MICHEL RIGO, AND MANON STIPULANTI

ABSTRACT. This paper is about counting the number of distinct (scattered) subwords occurring in a given word. More precisely, we consider the generalization of the Pascal triangle to binomial coefficients of words and the sequence $(S(n))_{n\geq 0}$ counting the number of positive entries on each row. By introducing a convenient tree structure, we provide a recurrence relation for $(S(n))_{n\geq 0}$. This leads to a connection with the 2-regular Stern–Brocot sequence and the sequence of denominators occurring in the Farey tree. Then we extend our construction to the Zeckendorf numeration system based on the Fibonacci sequence. Again our tree structure permits us to obtain recurrence relations for and the F-regularity of the corresponding sequence.

1. INTRODUCTION

A *finite word* is simply a finite sequence of letters belonging to a finite set called the *alphabet*. We recall from combinatorics on words that the binomial coefficient $\binom{u}{v}$ of two finite words u and v is the number of times v occurs as a subsequence of u (meaning as a "scattered" subword). As an example, consider the following binomial coefficient for two words over $\{0, 1\}$

$$\binom{101001}{101} = 6.$$

Indeed, if we index the letters of the first word $u_1 u_2 \cdots u_6 = 101001$, we have

$$u_1 u_2 u_3 = u_1 u_2 u_6 = u_1 u_4 u_6 = u_1 u_5 u_6 = u_3 u_4 u_6 = u_3 u_5 u_6 = 101.$$

Observe that this concept is a natural generalization of the binomial coefficients of integers. For a one-letter alphabet $\{a\}$, we have

$$(1) \qquad \binom{a^m}{a^n} = \binom{m}{n}, \quad \forall\, m, n \in \mathbb{N}$$

where a^m denotes the concatenation of m copies of the letter a. For more on these binomial coefficients, see, for instance, [26, Chap. 6]. There is a vast literature on the subject with applications in formal language theory (e.g., Parikh matrices, p-group languages or piecewise testable languages [21, 22]) and combinatorics on words (e.g., avoiding binomial repetitions [28]). One of the first combinatorial

questions was to determine when it is possible to uniquely reconstruct a word from some of its binomial coefficients; for instance, see [13].

The connection between the Pascal triangle and the Sierpiński gasket is well understood. Considering the intersection of the lattice $\mathbb{N}^2$ with the region $[0, 2^j] \times [0, 2^j]$, the first 2^j rows and columns of the usual Pascal triangle $(\binom{m}{n} \bmod 2)_{m,n<2^j}$ provide a coloring of this lattice. If we normalize this region by a homothety of ratio $1/2^j$, we get a sequence in $[0, 1] \times [0, 1]$ converging, for the Hausdorff distance, to the Sierpiński gasket when j tends to infinity. In [24] we extend this result for a generalized Pascal triangle defined below.

Definition 1. We let $\mathrm{rep}_2(n)$ denote the (greedy) base-2 expansion of $n \in \mathbb{N} \setminus \{0\}$ starting with 1 (these expansions correspond to expansions with most significant digit first). We set $\mathrm{rep}_2(0)$ to be the empty word denoted by ε. Let $L_2 = \{\varepsilon\} \cup 1\{0,1\}^*$ be the set of base-2 expansions of the integers. For all $i \geq 0$, we denote by w_i the ith word in L_2. Observe that $w_i = \mathrm{rep}_2(i)$ for $i \geq 0$.

To define an array P_2, we will consider all the words over $\{0, 1\}$ (starting with 1) and we order them by the radix order (i.e., first by length, then by the classical lexicographic ordering for words of the same length assuming $0 < 1$). For all $i, j \geq 0$, the element in the ith row and jth column of the array P_2 is defined as the binomial coefficient $\binom{w_i}{w_j}$ of the words w_i, w_j from Definition 1. This array is a generalized Pascal triangle that was introduced in [24] and its first few values are given in Table 1. These values correspond to the words $\varepsilon, 1, 10, 11, 100, 101, 110, 111$. In

		ε	1	10	11	100	101	110	111	S
$\mathrm{rep}_2(0)$	ε	**1**	0	0	0	0	0	0	0	1
$\mathrm{rep}_2(1)$	1	**1**	**1**	0	0	0	0	0	0	2
$\mathrm{rep}_2(2)$	10	1	1	1	0	0	0	0	0	3
$\mathrm{rep}_2(3)$	11	**1**	**2**	0	**1**	0	0	0	0	3
$\mathrm{rep}_2(4)$	100	1	1	2	0	1	0	0	0	4
$\mathrm{rep}_2(5)$	101	1	2	1	1	0	1	0	0	5
$\mathrm{rep}_2(6)$	110	1	2	2	1	0	0	1	0	5
$\mathrm{rep}_2(7)$	111	**1**	**3**	0	**3**	0	0	0	**1**	4

TABLE 1. The first few values in the generalized Pascal triangle P_2.

Table 1, the elements of the classical Pascal triangle are written in bold. A visual representation is given in Figure 1 where we have represented the positive values and a compressed version of the same figure. In Table 1, we also add an extra column in which we count, on each row of P_2, the number of positive binomial coefficients as explained in Definition 2.

Definition 2. We are interested in the sequence[1] $(S(n))_{n\geq 0}$ whose nth term, $n \geq 0$, is the number of non-zero elements in the nth row of P_2. Hence, the first few terms of this sequence are

$$1, 2, 3, 3, 4, 5, 5, 4, 5, 7, 8, 7, 7, 8, 7, 5, 6, 9, 11, 10, 11, 13, 12, 9, 9, 12, 13, 11, 10, \ldots.$$

---

[1]The sequence obtained by adding an extra 1 as a prefix of our sequence of interest matches exactly the sequence A007306 found in Sloane's On-Line Encyclopedia of Integer Sequences [33].
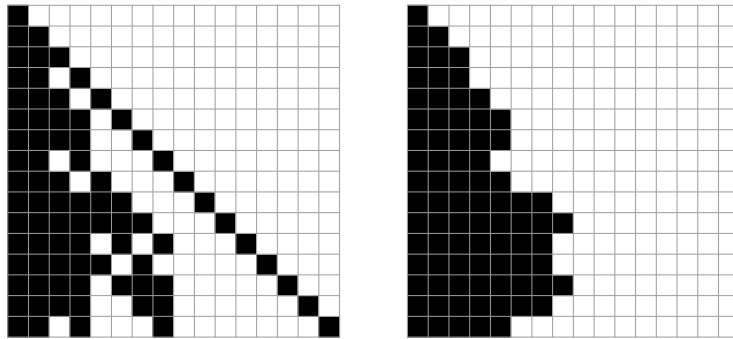
FIGURE 1. Positive values in the generalized Pascal triangle P_2 (on the left) and the compressed version (on the right).

Otherwise stated, for $n \geq 0$, we define

$$(2) \qquad S(n) := \# \left\{ v \in L_2 \ \middle| \ \binom{\mathrm{rep}_2(n)}{v} > 0 \right\}.$$

Observe that on a one-letter alphabet, i.e., for the classical Pascal triangle, the analogue sequence satisfies $S'(n) = n + 1$ for all $n \geq 0$. As we can see in Figure 2, the sequence $(S(n))_{n \geq 0}$ has a much more chaotic behavior.
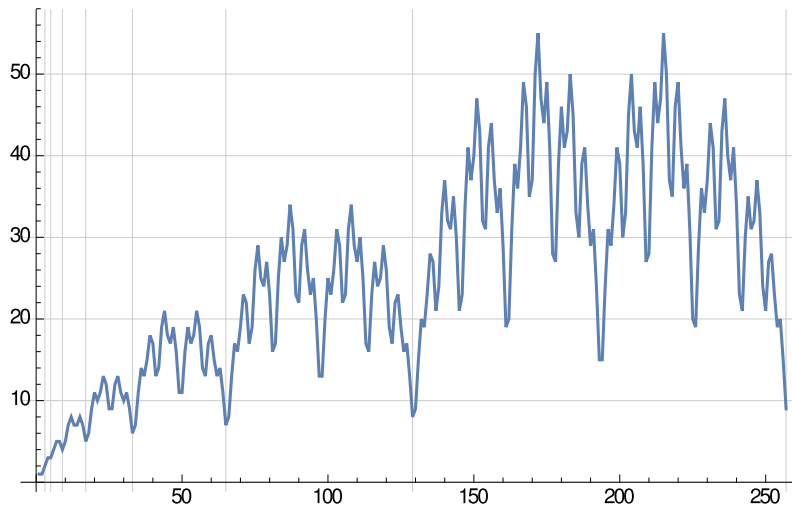

FIGURE 2. The sequence $(S(n))_{n \geq 0}$ in the interval $[0, 256]$.

However Figure 2 shares some similarities with the sequence studied in [27] and independently in [20] about the 2-abelian complexity of the Thue-Morse word and relative infinite words. For instance, we may observe a palindromic structure over each interval of the form $[2^n, 2^{n+1})$. This suggests that $(S(n))_{n \geq 0}$, as the Stern–Brocot sequence, should be 2-regular (see Section 5).

1.1. **Our contribution.** In Section 2, we introduce a convenient tree structure, called the trie of subwords, that permits us to easily count the number of subwords occurring in a given word. In Section 3, based on this notion, we obtain a recurrence relation satisfied by $(S(n))_{n \geq 0}$ and stated in Proposition 12. This result induces a relationship with the sequence A007306 and with the Stern–Brocot sequence $(SB(n))_{n \geq 0}$ A002487. In Section 4, we show that $S(n) = SB(2n + 1)$ for all $n \geq 0$. From the 2-regularity of the Stern–Brocot sequence [2, Example 7], one can immediately deduce that $(S(n))_{n \geq 0}$ is also 2-regular. In Section 5, using Proposition 12, we propose an alternative proof of that property. As a corollary, $S(n)$ can be computed by multiplying 2×2 matrices and the number of multiplications is proportional to $\log_2(n)$. Amongst the unbounded 2-regular sequences, the 2-synchronized sequences are those that can be "computed" with a finite automaton. In Section 6, we shortly discuss the fact that $(S(n))_{n \geq 0}$ is not 2-synchronized. In Section 7, we suggest that our techniques could be extended to base-k numeration systems for $k \geq 3$. This gives a motivation to study further such extensions of sequences related to the Stern–Brocot sequence.

In the last part of the paper, we consider a new variation of the Pascal triangle by reducing the entries to the admissible representations of integers in the Zeckendorf numeration system based on the Fibonacci sequence. Reconsidering our initial problem with this new array leads to a new sequence denoted by $(S_F(n))_{n \geq 0}$. Surprisingly, this sequence still behaves quite well. We get a recurrence relation similar to Proposition 12. Adapting the tries of subwords, we prove in Section 9 that $(S_F(n))_{n \geq 0}$ is F-regular (the definition will be recalled) in the sense of [32, 1] meaning that the $\mathbb{Z}$-module generated by the kernel adapted to the Zeckendorf numeration system is finitely generated. As a corollary, $S_F(n)$ can be computed by multiplying 2×2 matrices and the number of multiplications is proportional to $\log_\varphi(n)$ where φ is the golden ratio. Other sequences exhibiting this F-regularity can be found in [5, 14].

2. THE TRIE OF SUBWORDS

In [17], an automaton with multiplicities accepting exactly the subwords v of a given word u is presented. The number of accepting paths is exactly $\binom{u}{v}$. As mentioned in [22], it is an important problem to determine what the "best" data structures for reasoning with subwords are. Also see [8].

For our particular needs, we first introduce a convenient tree structure, where the nodes correspond to the subwords of a given word. This tree not only leads to a recurrence relation to compute $S(n)$ directly from $\mathrm{rep}_2(n)$ but can also be generalized to larger alphabets and to the Fibonacci case considered in the last part of the paper.

Definition 3. Let w be a finite word over $\{0, 1\}$. The language of its subwords is factorial, i.e., if xyz is a subword of w, then y is also a subword of w. Thus we may associate with w the *trie of its subwords* denoted by $\mathcal{T}(w)$. The root is ε, and if u and ua are two subwords of w with $a \in \{0, 1\}$, then ua is a child of u. This trie is also called prefix tree or radix tree. All successors of a node have a common prefix.

If we are interested in words and subwords belonging to a specific factorial language K, we only consider the part of $\mathcal{T}(w)$ that is a subtree of the tree associated with K. This subtree is denoted by $\mathcal{T}_K(w)$.

In the first part of this paper, we are interested in words and subwords belonging to L_2. We will thus consider the tree $\mathcal{T}_{L_2}(w)$. This means that in $\mathcal{T}(w)$ we will only consider the child 1 of the root ε.

Remark 4. Note that the number of nodes on level $\ell \geq 0$ in $\mathcal{T}_{L_2}(w)$ counts the number of subwords of length ℓ in L_2 occurring in w. In particular, the number of nodes of the trie $\mathcal{T}_{L_2}(\mathrm{rep}_2(n))$ is exactly $S(n)$ for all $n \geq 0$.

Example 5. In Figure 3, we have depicted the tree $\mathcal{T}_{L_2}(11001110)$ (the dashed lines and the subtrees T_ℓ, $\ell \in \{0, \ldots, 3\}$, will become clear with Definition 6 below). The word w is highlighted as the leftmost branch (it is more visual, if we do not use the convention that the left child of u is $u0$ and the right child of u is $u1$). The edge between u and its child $u0$ (resp., $u1$) is represented in gray (resp., black).
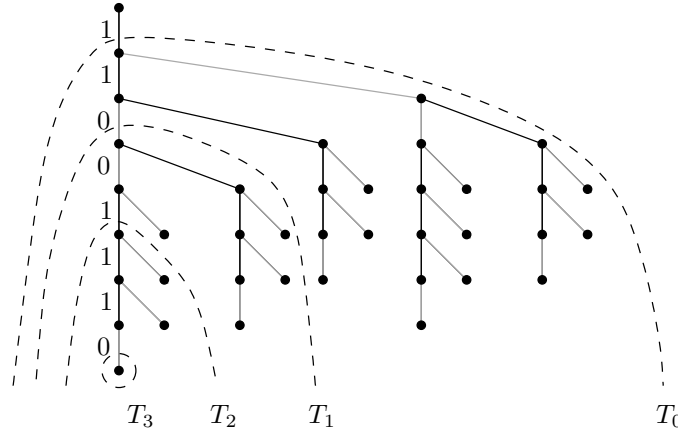

FIGURE 3. The trie $\mathcal{T}_{L_2}(11001110)$.

Since we are dealing with subwords of w, the tree $\mathcal{T}_{L_2}(w)$ has a particular structure that will be helpful to count the number of distinct subwords occurring in w. We now describe this structure which permits us to construct $\mathcal{T}_{L_2}(w)$ starting with a linear tree and proceeding bottom-up (see Example 8).

Definition 6. Each non-empty word w in L_2 is factorized into consecutive maximal blocks of ones and blocks of zeroes. It is of the form

$$(3) \qquad w = \underbrace{1^{n_1}}_{u_1}\underbrace{0^{n_2}}_{u_2}\underbrace{1^{n_3}}_{u_3}\underbrace{0^{n_4}}_{u_4} \cdots \underbrace{1^{n_{2j-1}}}_{u_{2j-1}}\underbrace{0^{n_{2j}}}_{u_{2j}}$$

with $j \geq 1$, $n_1, \ldots, n_{2j-1} \geq 1$ and $n_{2j} \geq 0$.

Let $M = M_w$ be such that $w = u_1 u_2 \cdots u_M$ where u_M is the last non-empty block of zeroes or ones. For every $\ell \in \{0, \ldots, M - 1\}$, the subtree of $\mathcal{T}_{L_2}(w)$ whose root is the node $u_1 \cdots u_\ell 1$ (resp., $u_1 \cdots u_\ell 0$) if ℓ is even (resp., odd) is denoted by T_ℓ (if $\ell = 0$, we set $u_1 \cdots u_\ell = \varepsilon$). For convenience (Corollary 9), we set T_M to be an empty tree with no node. Roughly speaking, we have a root of a new subtree T_ℓ for each new alternation of digits in w.

For the word 11001110 of Example 5, we have $M = 4$. In Figure 3, we have represented the trees $T_0, \ldots, T_3$ where T_3 is limited to a single node. Indeed, the number of nodes of T_{M-1} is n_M, which is equal to 1 in this example.

Since we are considering subwords of w, we immediately get the following.

Proposition 7. *Let w be a finite word in L_2. With the above notation about M and the subtrees T_ℓ, the tree $\mathcal{T}_{L_2}(w)$ has the following properties.*

- *Assume that $2 \le 2k < M$. For every $j \in \{0, \ldots, n_{2k} - 1\}$, the node of label $x = u_1 \cdots u_{2k-1} 0^j$ has two children $x0$ and $x1$. The node $x1$ is the root of a tree isomorphic to T_{2k}.*
- *Assume that $3 \le 2k + 1 < M$. Similarly, for every $j \in \{0, \ldots, n_{2k+1} - 1\}$, the node of label $x = u_1 \cdots u_{2k} 1^j$ has two children $x0$ and $x1$. The node $x0$ is the root of a tree isomorphic to T_{2k+1}.*
- *For every $j \in \{1, \ldots, n_1 - 1\}$, the node of label $x = 1^j$ has two children $x0$ and $x1$. The node $x0$ is the root of a tree isomorphic to T_1.*

As depicted in Figure 4, this proposition permits us to reconstruct the tree $\mathcal{T}_{L_2}(w)$ from w.

Example 8. In Figure 4, we show how to build $\mathcal{T}_{L_2}(11001110)$ in four steps. In this example, we have

$$11001110 = \underbrace{1^2}_{u_1} \underbrace{0^2}_{u_2} \underbrace{1^3}_{u_3} \underbrace{0^1}_{u_4}.$$

Hence the tree T_3 is limited to a single node because $n_4 = 1$. Actually, T_{M-1} is a linear tree with n_M nodes.

To reconstruct $\mathcal{T}_{L_2}(11001110)$, start with a linear tree corresponding to the word (it is depicted on the left in Figure 4). We proceed from the bottom of the tree. First, add a copy of T_3 to each node of the form $u_1 u_2 1^j$ for $j = 0, 1, 2$ (second picture).

Then consider the subtree T_2 whose root is the node $u_1 \cdots u_2 1$ and add a copy of it to each node of the form $u_1 0^j$ for $j = 0, 1$ (third picture).

Finally, consider the subtree T_1 whose root is the node $u_1 0$ and add a copy of it to the node 1 (picture on the bottom). Note that if $u_1 = 1^\ell$, we should add a copy of T_1 to each node of the form 1^j for $j = 1, \ldots, \ell - 1$ (if $\ell = 1$, then no copy of T_1 is added).

For the next statement, recall that T_M is an empty tree. If T is a tree, we let $\#T$ denote the number of nodes in T. This statement gives a recurrence relation to compute $S(n)$ directly from $\mathrm{rep}_2(n)$.

Corollary 9. *Let w be a finite word in L_2. With the above notation about M and the subtrees T_ℓ, the following holds. The number of nodes in T_{M-1} is n_M. For $j = M - 2, M - 3, \ldots, 0$, the number of nodes in T_j is given by*

$$\#T_j = n_{j+1}(\#T_{j+1} + 1) + \#T_{j+2}.$$

The number of subwords in L_2 of w is given by $1 + \#T_0$.

Proof. The idea is the same as the one developed in the previous example. Start with a linear tree labeled by w and add, with a bottom-up approach, all the possible subtrees given by Proposition 7: first possible copies of T_{M-1}, then T_{M-2}, ..., T_1. □

Remark 10. If we were interested in the number of subwords of w (not only those in L_2), we must add the node 0 which will be the root of a subtree equal to T_1. Thus, the total number of subwords occurring in w is $1 + \#T_0 + \#T_1$.
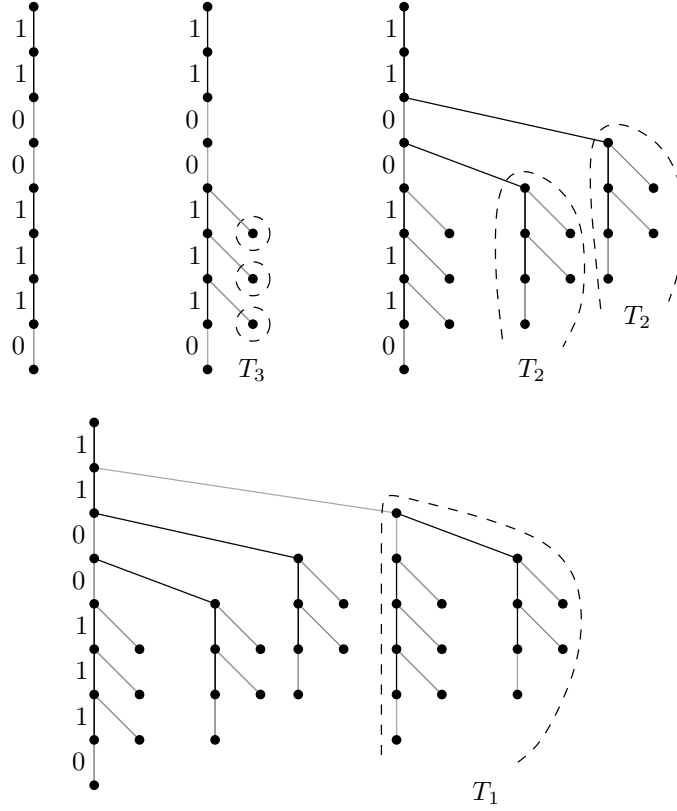
FIGURE 4. Bottom-up construction of $\mathcal{T}_{L_2}(11001110)$.

Example 11. For the word $w = 11001110$, $n_4 = 1$, $\#T_3 = 1$ and $\#T_4 = 0$. Thus, $\#T_2 = 3(1+1) + 0 = 6$, $\#T_1 = 2(6+1) + 1 = 15$, $\#T_0 = 2(15+1) + 6 = 38$. We get that the number of subwords of w in L_2 is 39 and since, $\mathrm{val}_2(11001110) = 206$, $S(206) = 39$. Moreover, the total number of subwords of 11001110 is $39 + 15 = 54$.

3. A RELATION SATISFIED BY THE SEQUENCE $(S(n))_{n \geq 0}$

Thanks to the trie of subwords introduced in the previous section, we collect several results about the number of words occurring as subwords of words with a prescribed form. The aim of this section is to prove the following result.

Proposition 12. *For all $\ell \geq 1$ and $0 \leq r < 2^\ell$, the sequence $(S(n))_{n \geq 0}$ given in Definition 2 satisfies $S(0) = 1$ and $S(1) = 2$ and*

$$(4) \qquad S(2^\ell + r) = \begin{cases} S(2^{\ell-1} + r) + S(r), & \text{if } 0 \leq r < 2^{\ell-1}; \\ S(2^{\ell+1} - r - 1), & \text{if } 2^{\ell-1} \leq r < 2^\ell. \end{cases}$$

This will permit us to show that the sequence $(S(n))_{n \geq 0}$ is the Farey sequence (see Section 4).

Lemma 13. *Let u be a word in $\{0,1\}^*$. Define $\underline{u}$ by replacing in u every 0 by 1 and every 1 by 0. Then*

$$\#\left\{v \in L_2 \mid \binom{1u}{v} > 0\right\} = \#\left\{v \in L_2 \mid \binom{1\underline{u}}{v} > 0\right\}.$$

In particular, this means that $S(2^\ell + r) = S(2^{\ell+1} - r - 1)$, if $2^{\ell-1} \le r < 2^\ell$.

Proof. It is enough to observe that the trees $\mathcal{T}_{L_2}(1u)$ and $\mathcal{T}_{L_2}(1\underline{u})$ are isomorphic. Each node of the form $1x$ in the first tree corresponds to the node $1\underline{x}$ in the second one and conversely.

For the special case, note that for every word z of length ℓ, there exists $r \in \{0, \ldots, 2^\ell - 1\}$ such that $\mathrm{rep}_2(2^\ell + r) = 1z$ and

$$\mathrm{val}_2(\underline{z}) = 2^\ell - r - 1 \in \{0 \ldots, 2^\ell - 1\}.$$

Hence, $1\underline{z} = \mathrm{rep}_2(2^{\ell+1} - r - 1)$. Using (2), we obtain the desired result. $\square$

Lemma 14. *Let u be a word in $\{0,1\}^*$. Then*

$$\#\left\{v \in L_2 \mid \binom{100u}{v} > 0\right\} = 2 \cdot \#\left\{v \in L_2 \mid \binom{10u}{v} > 0\right\} - \#\left\{v \in L_2 \mid \binom{1u}{v} > 0\right\}.$$

Proof. Our reasoning is again based on the structure of the trees. Assume first that u has no 1, then $u = 0^n$, $n \ge 0$. The tree $\mathcal{T}_{L_2}(1u)$ is linear and has $n + 2$ nodes, $\mathcal{T}_{L_2}(10u)$ has $n + 3$ nodes and $\mathcal{T}_{L_2}(100u)$ has $n + 4$ nodes. The formula holds.

Now assume that u has a 1. First observe that the subtree S of $\mathcal{T}_{L_2}(1u)$ with root 1 is equal to the subtree of $\mathcal{T}_{L_2}(10u)$ with root 10 and also to the subtree of $\mathcal{T}_{L_2}(100u)$ with root 100. Consider the shortest prefix of $1u$ of the form 10^r1. Let R be the subtree of $\mathcal{T}_{L_2}(1u)$ with root 10^r1. The subtree of $\mathcal{T}_{L_2}(10u)$ with root 11 is R. Similarly, $\mathcal{T}_{L_2}(100u)$ contains two copies of R: the subtrees of root 11 and 101. The situation is depicted in Figure 5 and the following formula holds: $3 + \#S + 2\#R = 2(2 + \#S + \#R) - (1 + \#S)$. $\square$
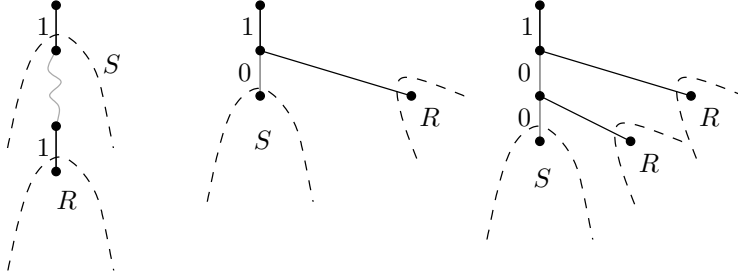
FIGURE 5. Structure of the trees.

Lemma 15. *Let u be a word in $\{0,1\}^*$. Then*

$$\#\left\{v \in L_2 \mid \binom{101u}{v} > 0\right\} = \#\left\{v \in L_2 \mid \binom{1u}{v} > 0\right\} + \#\left\{v \in L_2 \mid \binom{11u}{v} > 0\right\}.$$

Proof. The reasoning is similar to the one of the previous proof. If u is of the form 1^n, then $\mathcal{T}_{L_2}(101u)$ has $2n + 5$ nodes and $\mathcal{T}_{L_2}(1u)$ and $\mathcal{T}_{L_2}(11u)$ have respectively $n + 2$ and $n + 3$ nodes. If u has a 0, then consider the shortest prefix of $101u$ of the

form 101^r0. Let S be the subtree of $\mathcal{T}_{L_2}(101u)$ with root 101 and R be the subtree with root 101^r0. The tree $\mathcal{T}_{L_2}(101u)$ (resp., $\mathcal{T}_{L_2}(1u)$, $\mathcal{T}_{L_2}(11u)$) has $3 + 2\#S + \#R$ (resp., $1 + \#S$, $2 + \#S + \#R$) nodes. $\qquad\square$

Proof of Proposition 12. Consider some integers ℓ and r with $\ell \geq 1$ and $0 \leq r < 2^\ell$. If $2^{\ell-1} \leq r < 2^\ell$, the equality $S(2^\ell + r) = S(2^{\ell+1} - r - 1)$ directly follows from Lemma 13.

So let us suppose $0 \leq r < 2^{\ell-1}$. We proceed by induction on ℓ. The case $\ell = 1$ is easily checked. Let us suppose $\ell \geq 2$. By definition we have

$$S(2^\ell + r) = \#\left\{ v \in L_2 \mid \binom{\mathrm{rep}_2(2^\ell + r)}{v} > 0 \right\}.$$

Since $r < 2^{\ell-1}$, we have $\mathrm{rep}_2(2^\ell + r) = 10u$ for some word $u \in \{0,1\}^*$ of length $\ell - 1$ such that $r = \mathrm{val}_2(u)$. We consider two cases depending on the first letter occurring in u.

If $u \in 0\{0,1\}^*$, then $r = \mathrm{val}_2(u) \leq 2^{\ell-2} - 1$ and we deduce from Lemma 14 that

$$S(2^\ell + r) = 2S(2^{\ell-1} + r) - S(2^{\ell-2} + r).$$

The proof is complete after using the induction hypothesis twice:

$$
\begin{aligned}
S(2^\ell + r) &= 2(S(2^{\ell-2} + r) + S(r)) - S(2^{\ell-2} + r) \\
&= S(2^{\ell-2} + r) + S(r) + S(r) \\
&= S(2^{\ell-1} + r) + S(r).
\end{aligned}
$$

If $u = 1u'$ for some $u' \in \{0,1\}^*$, then $\mathrm{val}_2(101u') = 2^\ell + r$, $\mathrm{val}_2(1u') = \mathrm{val}_2(u) = r$ and $\mathrm{val}_2(11u') = 2^{\ell-1} + r$. We deduce from Lemma 15 applied to u' that

$$S(2^\ell + r) = S(r) + S(2^{\ell-1} + r),$$

which finishes the proof. $\qquad\square$

4. The sequence A007306, Farey tree and Stern–Brocot sequence

Plugging in the first few terms of $(S(n))_{n\geq 0}$ in Sloane's On-Line Encyclopedia of Integer Sequences [33], it seems to be a shifted version of the sequence A007306 of the denominators occurring in the Farey tree (left subtree of the full Stern–Brocot tree), which contains every (reduced) positive rational less than 1 exactly once. Many papers deal with this tree; for instance, see [4, 18]. The *Farey tree* is an infinite binary tree made up of mediants. Given two reduced fractions $\frac{a}{b}$ and $\frac{c}{d}$, with $a, b, c, d \in \mathbb{N}$, their *mediant* is the fraction $\frac{a}{b} \oplus \frac{c}{d} = \frac{a+c}{b+d}$. This operation is known as *child's addition*. Observe that for all $\frac{a}{b}$, $\frac{c}{d}$ with $\frac{a}{b} < \frac{c}{d}$, we have $\frac{a}{b} < \frac{a}{b} \oplus \frac{c}{d} < \frac{c}{d}$.

Definition 16. Starting from the fractions $\frac{0}{1}$ and $\frac{1}{1}$, the *Farey tree* is the infinite tree defined as follows:

  (1) the set of nodes is partitioned into levels indexed by $\mathbb{N}$;
  (2) the level 0 consists in $\{\frac{0}{1}, \frac{1}{1}\}$;
  (3) the level 1 consists in $\{\frac{1}{2}\}$. The node $\frac{1}{2}$ is the only one with two parents, which are $\frac{0}{1}$ and $\frac{1}{1}$;

(4) for each $n \geq 2$, the level n consists in the children of the nodes of vertices in level $n-1$. For each node $\frac{a}{b}$ of level $n-1$, we define

$$\text{Left}\left(\frac{a}{b}\right) = \max\left\{\frac{e}{f} \mid \text{level}\left(\frac{e}{f}\right) < n-1 \text{ and } \frac{e}{f} < \frac{a}{b}\right\};$$

$$\text{Right}\left(\frac{a}{b}\right) = \min\left\{\frac{e}{f} \mid \text{level}\left(\frac{e}{f}\right) < n-1 \text{ and } \frac{e}{f} > \frac{a}{b}\right\}.$$

The left and right children of $\frac{a}{b}$ are respectively $\frac{a}{b} \oplus \text{Left}\left(\frac{a}{b}\right)$ and $\frac{a}{b} \oplus \text{Right}\left(\frac{a}{b}\right)$.
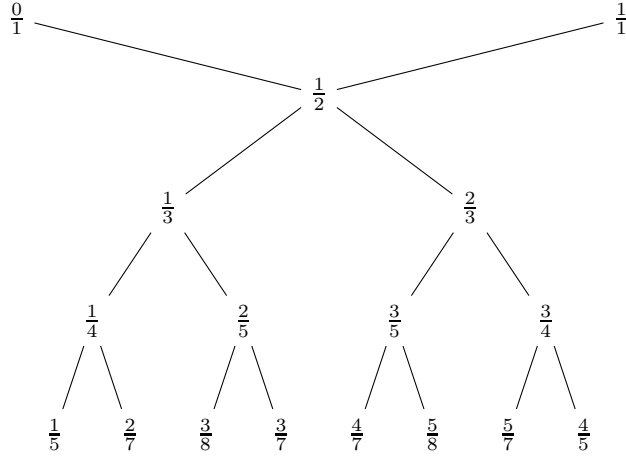

FIGURE 6. The first levels of the Farey tree.

Reading the denominators in Figure 6 level-by-level, then from left to right, i.e., conducting a breadth-first traversal of the tree, we obtain the first few terms of the sequence $(S(n))_{n \geq 0}$ if we drop the second denominator 1. If a branch to the left (resp., right) corresponds to 0 (resp., 1), then with every node $\frac{a}{b}$ (except $\frac{1}{1}$) is associated a unique path from $\frac{0}{1}$ to $\frac{a}{b}$ of label u. If $n = \text{val}_2(u)$, then we will show that $b = S(n)$. For instance, $\frac{3}{7}$ corresponds to the path of label $1011 = \text{rep}_2(11)$ and $S(11) = 7$. Conversely, given an integer $n \geq 0$, we let $D(n)$ be the denominator of the node reached from $\frac{0}{1}$ using the path of label $\text{rep}_2(n)$. The sequence $(D(n))_{n \geq 0}$ is indexed by A007306 in [33].

We show that the sequence $(D(n))_{n \geq 0}$ satisfies $D(n) = S(n)$ for all n. From the definition of the tree, for all $\ell \geq 1$, $u \in \{0,1\}^*$, it directly follows that

$$D(\text{val}_2(u10^\ell)) = D(\text{val}_2(u10^{\ell-1})) + D(\text{val}_2(u))$$

and

$$D(\text{val}_2(u01^\ell)) = D(\text{val}_2(u01^{\ell-1})) + D(\text{val}_2(u)).$$

Otherwise stated, if $n = 2^k + r$ with $0 \leq r < 2^k$, we have

$$D(n) = \begin{cases} D(n/2) + D((n-2^\ell)/2^{\ell+1}), & \text{if } \text{rep}_2(r) \in \{0,1\}^*10^\ell; \\ D((n-1)/2) + D((n-2^\ell+1)/2^{\ell+1}), & \text{if } \text{rep}_2(r) \in \{0,1\}^*01^\ell. \end{cases}$$

We show by induction that the sequence $(S(n))_{n \geq 0}$ satisfies the same formulas, i.e.,

$$S(n) = \begin{cases} S(n/2) + S((n - 2^\ell)/2^{\ell+1}), & \text{if } \operatorname{rep}_2(r) \in \{0,1\}^* 10^\ell; \\ S((n-1)/2) + S((n - 2^\ell + 1)/2^{\ell+1}), & \text{if } \operatorname{rep}_2(r) \in \{0,1\}^* 01^\ell. \end{cases}$$

Note that a direct inspection shows that $S(n) = D(n), 0 \leq n \leq 3$. For the induction step, we write $n = 2^k + r$ with $k \geq 2$ and $0 \leq r < 2^k$. Using (4), we then consider four cases, depending on whether $0 \leq r < 2^{k-1}$ or $2^{k-1} \leq r < 2^k$ and whether $\operatorname{rep}_2(r)$ has a suffix consisting of zeroes or of ones. We only give the proof for the case $0 \leq r < 2^{k-1}$ with $\operatorname{rep}_2(r) \in \{0,1\}^* 10^\ell$, $\ell \geq 1$, the other ones being similar. We have

$$\begin{aligned} S(n) &= S(2^{k-1} + r) + S(r) & \text{(by (4))} \\ &= S((2^{k-1} + r)/2) + S((2^{k-1} + r - 2^\ell)/2^{\ell+1}) \\ &\quad + S(r/2) + S((r - 2^\ell)/2^{\ell+1}) & \text{(by induction hypothesis)} \\ &= S(2^{k-1} + r/2) + S(2^{k-\ell-1} + r/2^{\ell+1} + 1/2) & \text{(by (4))} \\ &= S(n/2) + S((n - 2^\ell)/2^{\ell+1}). \end{aligned}$$

Remark 17. It is a folklore fact that the sum of the denominators of the elements of the kth level (with $k \geq 1$) in the Farey tree is equal to $2 \cdot 3^{k-1}$ and is 1 if we only consider the denominator $D(0)$. Or similarly, that the sum of the denominators of the levels 1 to k is equal to $3^k - 1$ (or 3^k if we add the denominator $D(0)$ on the 0th level). Using (2), we observe that $\sum_{i=0}^{2^n - 1} S(i)$ is the number of pairs of words in L_n having a positive binomial coefficient. In [24], we showed that this number is equal to 3^n, giving an alternative proof of the above mentioned fact about the Farey tree.

The *Stern–Brocot sequence* $(SB(n))_{n \geq 0}$ is defined by $SB(0) = 0, SB(1) = 1$ and, for all $n \geq 1$,

$$SB(2n) = SB(n), \quad SB(2n+1) = SB(n) + SB(n+1).$$

It is well known that $D(n) = SB(2n+1)$. Hence the sequence $(S(n))_{n \geq 0}$ satisfies

$$(5) \qquad\qquad S(n) = SB(2n+1) \quad \text{for all } n \geq 0.$$

Remark 18. In [9], it is shown that the nth Stern–Brocot value $SB(n)$ is equal to the number of times words of the form $v \in 1(01)^*$ occur as scattered subwords of the binary expansion of n. This result is different from the one obtained here because the form of the subwords is fixed. In [11], the authors give a way to build the sequence $(SB(n))_{n \geq 0}$ using occurrences of words occurring in the base-2 expansions of positive integers.

5. THE SEQUENCE $(S(n))_{n \geq 0}$ IS 2-REGULAR

The k-regularity of a sequence provides interesting structural information about it. For instance, we get matrices to compute its n-th term in a number of steps proportional to $\log_k(n)$. First we recall the notions of k-kernel and k-regular sequence. Since $(SB(n))_{n \geq 0}$ is 2-regular [2], one can immediately deduce that the sequence $(S(n))_{n \geq 0}$ is 2-regular from (5). Nevertheless, we provide an alternative proof because we have in mind extensions to other numeration systems; see Section 7 and Section 8.

Definition 19. Let $k \geq 2$ be an integer. The k-*kernel* of a sequence $s = (s(n))_{n \geq 0}$ is the set

$$\mathcal{K}_k(s) = \{(s(k^i n + j))_{n \geq 0}|\ i \geq 0 \text{ and } 0 \leq j < k^i\}.$$

One characterization of k-automatic sequences is that their k-kernels are finite; see [15] or [3]. For instance, the 2-kernel $\mathcal{K}_2(\mathbf{t})$ of the Thue–Morse word $\mathbf{t} = 01101001 \cdots$ contains exactly two elements, namely $\mathbf{t}$ and $\underline{\mathbf{t}}$. Unbounded sequences (i.e., taking infinitely many integer values) are also of interest but clearly, their k-kernels are infinite. To handle such sequences, one introduces the following definition in the sense of Allouche and Shallit [2]. Also see [6].

Definition 20. Let $k \geq 2$ be an integer. A sequence $s = (s(n))_{n \geq 0}$ is k-*regular* if $\langle \mathcal{K}_k(s) \rangle$ is a finitely-generated $\mathbb{Z}$-module, i.e., there exists a finite number of sequences $(t_1(n))_{n \geq 0}, \ldots, (t_\ell(n))_{n \geq 0}$ such that every sequence in the $\mathbb{Z}$-module generated by the k-kernel $\mathcal{K}_k(s)$ is a $\mathbb{Z}$-linear combination of the t_r's. Otherwise stated, for all $i \geq 0$ and for all $j \in \{0, \ldots, k^i - 1\}$, there exist integers $c_1, \ldots, c_\ell$ such that

$$\forall n \geq 0, \quad s(k^i n + j) = \sum_{r=1}^{\ell} c_r\, t_r(n).$$

As an easy example of 2-regular (unbounded) sequence, consider the Stern–Brocot sequence. The $\mathbb{Z}$-module generated by the 2-kernel of $(SB(n))_{n \geq 0}$ is simply generated by the sequence itself and the shifted sequence $(SB(n+1))_{n \geq 0}$.

Theorem 21. *The sequence $(S(n))_{n \geq 0}$ satisfies, for all $n \geq 0$,*

$$
\begin{aligned}
S(2n+1) &= 3\,S(n) - S(2n) \\
S(4n) &= 2\,S(2n) - S(n) \\
S(4n+2) &= 4\,S(n) - S(2n).
\end{aligned}
$$

In particular, $(S(n))_{n \geq 0}$ is 2-regular.

Proof. We first prove $S(2n+1) + S(2n) = 3S(n)$. We proceed by induction on n. It can be checked by hand that the result holds true for $n \leq 1$. Thus consider $n > 1$ and suppose that the relation holds true for all $m < n$. We write $n = 2^\ell + r$ with $\ell \geq 1$ and $0 \leq r < 2^\ell$. We divide the proof in two according to the position of r inside the interval $[0, 2^\ell)$.

(a) If $0 \leq r < 2^{\ell-1}$, we get

$$
\begin{aligned}
S(2n+1) + S(2n) &= S(2^{\ell+1} + 2r + 1) + S(2^{\ell+1} + 2r) \\
&= S(2^\ell + 2r + 1) + S(2r + 1) + S(2^\ell + 2r) + S(2r) && \text{(by (4))} \\
&= 3S(2^{\ell-1} + r) + 3S(r) && \text{(by induction hypothesis)} \\
&= 3S(n) && \text{(by (4))}.
\end{aligned}
$$

(b) If $2^{\ell-1} \leq r < 2^\ell$, we get

$$
\begin{aligned}
S(2n+1) + S(2n) &= S(2^{\ell+1} + 2r + 1) + S(2^{\ell+1} + 2r) \\
&= S(2^{\ell+2} - 2r - 2) + S(2^{\ell+2} - 2r - 1) && \text{(by (4))} \\
&= 3S(2^{\ell+1} - r - 1) && \text{(by induction hypothesis)} \\
&= 3S(n) && \text{(by (4))}.
\end{aligned}
$$

Now we prove the remaining two relations. We again proceed by induction on n. It can be checked by hand that the result holds true for $n \leq 1$. Thus consider $n > 1$ and suppose that the two relations hold true for all $m < n$.

(1) Let us prove $S(4n) = 2S(2n) - S(n)$. We write $n = 2^\ell + r$ with $\ell \geq 1$ and $0 \leq r < 2^\ell$.
   (a) If $0 \leq r < 2^{\ell-1}$, we get

$$
\begin{aligned}
S(4n) &= S(2^{\ell+2} + 4r) \\
&= S(2^{\ell+1} + 4r) + S(4r) && \text{(by (4))} \\
&= 2S(2^\ell + 2r) - S(2^{\ell-1} + r) + 2S(2r) - S(r) && \text{(by induction hypothesis)} \\
&= 2S(2n) - S(n) && \text{(by (4)).}
\end{aligned}
$$

   (b) If $2^{\ell-1} \leq r < 2^\ell$, we get

$$
\begin{aligned}
S(4n) &= S(2^{\ell+2} + 4r) \\
&= S(2^{\ell+3} - 4r - 1) && \text{(by (4))} \\
&= 3S(2^{\ell+2} - 2r - 1) - S(2^{\ell+3} - 4r - 2) && \text{(using the first relation)} \\
&= 3S(2n) - S(2^{\ell+3} - 4(r+1) + 2) && \text{(by (4))} \\
&= 3S(2n) - 4S(2^{\ell+1} - r - 1) + S(2^{\ell+2} - 2r - 2) && \text{(by induction hypothesis)} \\
&= 3S(2n) - 4S(n) + S(2n+1) && \text{(by (4))} \\
&= 2S(2n) - S(n) && \text{(using the first relation).}
\end{aligned}
$$

(2) Finally, let us prove $S(4n+2) = 4S(n) - S(2n)$. Let us write $n = 2^\ell + r$ with $\ell \geq 1$ and $0 \leq r < 2^\ell$.
   (a) If $0 \leq r < 2^{\ell-1}$, we get

$$
\begin{aligned}
S(4n+2) &= S(2^{\ell+2} + 4r + 2) \\
&= S(2^{\ell+1} + 4r + 2) + S(4r + 2) && \text{(by (4))} \\
&= 4S(2^{\ell-1} + r) - S(2^\ell + 2r) + 4S(r) - S(2r) && \text{(by induction hypothesis)} \\
&= 4S(n) - S(2n) && \text{(by (4)).}
\end{aligned}
$$

   (b) If $2^{\ell-1} \leq r < 2^\ell$, we get

$$
\begin{aligned}
S(4n+2) &= S(2^{\ell+2} + 4r + 2) \\
&= S(2^{\ell+3} - 4r - 3) && \text{(by (4))} \\
&= 3S(2^{\ell+2} - 2r - 2) - S(2^{\ell+3} - 4r - 4) && \text{(using the first relation)} \\
&= 3S(2n+1) - S(2^{\ell+3} - 4(r+1)) && \text{(by (4))} \\
&= 3S(2n+1) - 2S(2^{\ell+2} - 2r - 2) + S(2^{\ell+1} - r - 1) && \text{(by induction hypothesis)} \\
&= 3S(2n+1) - 2S(2n+1) + S(n) && \text{(by (4))} \\
&= 4S(n) - S(2n) && \text{(using the first relation).}
\end{aligned}
$$

To finish the proof, observe that the $\mathbb{Z}$-module $\langle \mathcal{K}_2(S) \rangle$ is finitely-generated: a choice of generators is $(S(n))_{n \geq 0}$ and $(S(2n))_{n \geq 0}$. $\qquad\square$

If a sequence is k-regular, then its nth term can be obtained by multiplying some matrices and the length of this product is proportional to $\log_k(n)$. In our situation, we will consider products of square matrices of size 2, [2], [3, Theorem 16.1.3]. Observe that due to Equation (5), other matrices can be derived from the 2-regularity of $(SB(n))_{n \geq 0}$.

Corollary 22. *For all $n \geq 0$, let*

$$
(6) \qquad V(n) = \begin{pmatrix} S(n) \\ S(2n) \end{pmatrix}.
$$

Consider the matrix-valued morphism $\mu : \{0,1\}^* \to \mathbb{Z}_2^2$ defined by

$$
\mu(0) = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}, \quad \mu(1) = \begin{pmatrix} 3 & -1 \\ 4 & -1 \end{pmatrix}.
$$

Then $V(2n+r) = \mu(r)V(n)$ for all $r \in \{0,1\}$ and $n \geq 0$. Consequently, if $\mathrm{rep}_2(n) = c_r \cdots c_0$, then

$$S(n) = \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(c_0) \cdots \mu(c_r) V(0).$$

Proof. Thanks to Theorem 21, we directly have

$$V(2n) = \begin{pmatrix} S(2n) \\ S(4n) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} S(n) \\ S(2n) \end{pmatrix}$$

and

$$V(2n+1) = \begin{pmatrix} S(2n+1) \\ S(4n+2) \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} S(n) \\ S(2n) \end{pmatrix}$$

for all $n \geq 0$. Now let $r = \sum_{i=0}^{\ell-1} r_i \, 2^i$. Then the word $r_{\ell-1} \cdots r_0$ is the expression of r in base 2 possibly with leading zeroes. By induction, we can show that

$$(7) \qquad\qquad V(2^\ell m + r) = \mu(r_0 \cdots r_{\ell-1})V(m)$$

for all $m \in \mathbb{N}$. Now let $n \geq 2$. Then there exist $\ell \geq 1$ and $r \in \{0, \ldots, 2^\ell - 1\}$ such that $n = 2^\ell + r$. Let $r_{\ell-1} \cdots r_0$ be the expression of r in base 2 possibly with leading zeroes. Using (7) and the fact that $V(1) = \mu(1)V(0)$, we get

$$V(n) = \mu(r_0 \cdots r_{\ell-1})V(1) = \mu(r_0 \cdots r_{\ell-1}1)V(0) = \mu((\mathrm{rep}_2(n))^R)V(0)$$

where u^R is the reversal of the word u. Note that the previous equality also holds for $n \in \{0,1\}$. In particular, we have the following equality

$$S(n) = \begin{pmatrix} 1 & 0 \end{pmatrix} \mu((\mathrm{rep}_2(n))^R V(0)$$

for all $n \in \mathbb{N}$. $\square$

6. THE SEQUENCE $(S(n))_{n \geq 0}$ IS NOT 2-SYNCHRONIZED

The class of k-synchronized sequences is an intermediate between the classes of k-automatic sequences and k-regular sequences. Every k-synchronized sequence is k-regular but the converse does not hold. Moreover, every k-synchronized sequence taking finitely many values is k-automatic. Roughly, a sequence $(s(n))_{n \geq 0}$ is k-synchronized if there exists a finite automaton accepting the pairs of base-k expansions of n and $s(n)$, as stated in Definition 23. These sequences were first introduced in [10]. As an example, it is proved in [19] that if an infinite word $\mathbf{w}$ is k-automatic, then its factor complexity function is k-synchronized.

Definition 23. Let $k \geq 2$ be an integer. The map rep_k is extended to $\mathbb{N} \times \mathbb{N}$ as follows. For all $m, n \in \mathbb{N}$,

$$\mathrm{rep}_k(m,n) = \left(0^{M-|\mathrm{rep}_k(m)|}\mathrm{rep}_k(m), 0^{M-|\mathrm{rep}_k(n)|}\mathrm{rep}_k(n) \right)$$

where $M = \max\{|\mathrm{rep}_k(m)|, |\mathrm{rep}_k(n)|\}$. The idea is that the shortest word is padded with leading zeroes to get two words of the same length.

A sequence $(s(n))_{n \geq 0}$ of integers is said to be *k-synchronized* if the language $\{\mathrm{rep}_k(n, s(n)) \mid n \in \mathbb{N}\}$ is accepted by some finite automaton reading pairs of digits.

We present two independent proofs of the next result.

Proposition 24. *The sequence $(S(n))_{n \geq 0}$ is not 2-synchronized.*

Proof. Proceed by contradiction and suppose that there is a deterministic k-state automaton that accepts exactly the language $\{\mathrm{rep}_2(n, S(n)) \mid n \in \mathbb{N}\}$. Note that for all $\ell \geq 0$, $S(2^\ell) = \ell + 2$ using (2). Consider an integer ℓ such that $\ell - \lceil \log_2(\ell+2) \rceil > k + 1$. Then we have $\mathrm{rep}_2(2^\ell, \ell + 2) = (10^\ell, 0^{k+2}u)$ for some word $u \in \{0,1\}^*$ of length $\ell - k - 1$ and such that $\mathrm{val}_2(u) = \ell + 2$. Let $(q_0, q_1, \ldots, q_{\ell+2})$ be the path in the automaton starting in the initial state q_0 and whose label is $\mathrm{rep}_2(2^\ell, \ell+2)$. Since $\mathrm{rep}_2(2^\ell, \ell + 2)$ is accepted by the automaton, the state $q_{\ell+2}$ is an accepting state. As the automaton has k states, there exist $1 \leq i < j \leq k + 2$ such that $q_i = q_j$. By choice of ℓ, there is a path from q_i to q_j whose label is $(0^{j-i}, 0^{j-i})$. Thus the pair of words $(10^{\ell+j-i}, 0^{k+2+j-i}u)$ is accepted by the automaton. However, we have $S(\mathrm{val}_2(10^{\ell+j-i})) = \ell + j - i + 2 \neq \mathrm{val}_2(0^{k+2+j-i}u) = \ell + 2$, which is a contradiction. $\qquad\square$

The fact that the sequence $(S(n))_{n \geq 0}$ is not 2-synchronized also follows from the next result.

Lemma 25. [31, Lemma 4] *If $(f(n))_{n \geq 0}$ is a k-synchronized sequence, and $f \neq O(1)$, then there exists a constant $c > 0$ such that $f(n) \geq cn$ infinitely often.*

We can make use of the growth order of the Stern–Brocot sequence. One can numerically estimate that the joint spectral radius ρ of the matrices $\mu(0)$ and $\mu(1)$ is between 1.61 and 1.71. Since, for all $n \geq 0$, $S(n)$ can be computed by multiplying those matrices and the number of multiplications is proportional to $\log_2(n)$, it follows that the growth rate of $S(n)$ cannot be higher than $\rho^{\log_2(n)} = n^{\log_2(\rho)}$ multiplied by a constant. By Lemma 25, the sequence $(S(n))_{n \geq 0}$ cannot be 2-synchronized.

7. Extension to a larger alphabet

In this section, we investigate the case of the base-k numeration system for $k \geq 2$. We are able to extend our main tool that are the tries of subwords defined in Section 2. However, it is not straightforward to deduce an analogue of Proposition 12 from Corollary 29. Nevertheless, numerical computations suggest that the analogue of the sequence $(S(n))_{n \geq 0}$ in the base-k case seems to be k-regular. In [25], we prove an exact behavior for the summary function of $(S(n))_{n \geq 0}$ and conjecture a similar behavior for larger bases. General asymptotic estimates for summary function of k-regular sequences are provided by Dumas in [12].

For instance, if $(S_3(n))_{n \geq 0}$ denotes the analogue of the sequence $(S(n))_{n \geq 0}$ in the base-3 case, then the first few terms of this sequence are

$$1, 2, 2, 3, 3, 4, 3, 4, 3, 4, 5, 6, 5, 4, 6, 7, 7, 6, 4, 6, 5, 7, 6, 7, 5, 6, 4, 5, 7, 8, 8, 7, 10, \ldots.$$

We conjecture the following two results that are the analogues of Proposition 12 and Theorem 21. For all $\ell \geq 1$ and $0 \leq r < 3^\ell$, the sequence $(S_3(n))_{n \geq 0}$ satisfies $S_3(0) = 1$, $S_3(1) = S_3(2) = 2$,

$$S_3(3^\ell + r) = \begin{cases} S_3(3^{\ell-1} + r) + S_3(r), & \text{if } 0 \leq r < 3^{\ell-1}; \\ 2S_3(r) - S_3(r - 3^{\ell-1}), & \text{if } 3^{\ell-1} \leq r < 2 \cdot 3^{\ell-1}; \\ 2S_3(r) + S_3(r - 3^{\ell-1}) - 2S_3(r - 2 \cdot 3^{\ell-1}), & \text{if } 2 \cdot 3^{\ell-1} \leq r < 3^\ell; \end{cases}$$

and

$$S_3(2 \cdot 3^\ell + r) = \begin{cases} S_3(2 \cdot 3^{\ell-1} + r) + S_3(r), & \text{if } 0 \le r < 3^{\ell-1}; \\ S_3(r + 3^{\ell-1}) + 2S_3(r) - 2S_3(r - 3^{\ell-1}), & \text{if } 3^{\ell-1} \le r < 2 \cdot 3^{\ell-1}; \\ 2S_3(r) - S_3(r - 2 \cdot 3^{\ell-1}), & \text{if } 2 \cdot 3^{\ell-1} \le r < 3^\ell. \end{cases}$$

Observe that we divide the statement into two cases depending on the first letter of the base-3 expansion of the integers. Moreover, there exists a partial palindromic structure inside of the sequence $(S_3(n))_{n \ge 0}$ that can be stated as follows: for all $\ell \ge 1$ and $0 \le r < 3^\ell$,

$$S_3(2 \cdot 3^\ell + r) = S_3(2 \cdot 3^\ell + 3^\ell - r - 1).$$

Conjecture 26. *The sequence $(S_3(n))_{n \ge 0}$ satisfies, for all $n \ge 0$,*

$$\begin{aligned} S_3(3n+2) &= 5\,S_3(n) - S_3(3n) - S_3(3n+1) \\ S_3(9n) &= -S_3(n) + 2\,S_3(3n) \\ S_3(9n+1) &= -2\,S_3(n) + 2\,S_3(3n) + S_3(3n+1) \\ S_3(9n+3) &= -2\,S_3(n) + S_3(3n) + 2\,S_3(3n+1) \\ S_3(9n+4) &= -S_3(n) + 2\,S_3(3n+1) \\ S_3(9n+6) &= 8\,S_3(n) - S_3(3n) - 2\,S_3(3n+1) \\ S_3(9n+7) &= 8\,S_3(n) - 2\,S_3(3n) - S_3(3n+1). \end{aligned}$$

In particular, $(S_3(n))_{n \ge 0}$ is 3-regular.

For each $k \ge 2$, we let L_k denote the language $\{1, \ldots, k-1\}\{0, \ldots, k-1\}^* \cup \{\varepsilon\}$. Let us introduce some notation. As in Section 2, for each non-empty word $w \in L_k$, we consider an equivalent factorization of w into maximal blocks of letters of the form

$$w = a_1^{n_1} \cdots a_M^{n_M},$$

with $n_\ell \ge 1$ for all ℓ. For each $\ell \in \{0, \ldots, M-1\}$, we consider the subtree T_ℓ of $\mathcal{T}_{L_k}(w)$ whose root is the node $a_1^{n_1} \cdots a_\ell^{n_\ell} a_{\ell+1}$. We again set T_M to be an empty tree with no node.

For each $\ell \in \{0, \ldots, M-1\}$, let $\text{Alph}(\ell)$ denote the set of letters occurring in $a_{\ell+1} \cdots a_M$. Then for each letter $b \in \text{Alph}(\ell)$, we let $j(b, \ell)$ denote the first index greater than ℓ such that $a_{j(b,\ell)} = b$. We obtain the analogue of Proposition 7.

Proposition 27. *Let w be a finite word in L_k. With the above notation about M and the subtrees T_ℓ, the tree $\mathcal{T}_{L_k}(w)$ has the following properties.*

(1) *The node of label ε has $\#(\text{Alph}(0) \backslash \{0\})$ children that are b for $b \in \text{Alph}(0) \backslash \{0\}$. Each child b is the root of a tree isomorphic $T_{j(b,0)-1}$.*

(2) *For each $\ell \in \{0, \ldots, M-1\}$ and each $i \in \{0, \ldots, n_{\ell+1} - 1\}$ with $(\ell, i) \ne (0, 0)$, the node of label $x = a_1^{n_1} \cdots a_\ell^{n_\ell} a_{\ell+1}^i$ has $\#(\text{Alph}(\ell))$ children that are xb for $b \in \text{Alph}(\ell)$. Each child xb with $b \ne a_{\ell+1}$ is the root of a tree isomorphic to $T_{j(b,\ell)-1}$.*

Example 28. Let $w = 22000112 \in L_3$. The tree $\mathcal{T}_{L_3}(w)$ is depicted in Figure 7. We follow the same lines as for Example 5. We use three different colors to represent the letters $0, 1, 2$.

With a similar proof, we then get the following analogue of Corollary 9.

FIGURE 7. The trie $\mathcal{T}_{L_2}(11001110)$.

Corollary 29. *Let w be a finite word in L_k. With the above notation about M and the subtrees T_ℓ, the following holds. The number of nodes in T_{M-1} is n_M. For $i = M-2, M-3, \ldots, 0$, the number of nodes in T_i is given by*

$$\#T_i = n_{i+1}\left(1 + \sum_{\substack{b \in \mathrm{Alph}(i+1) \\ b \neq a_{i+1}}} \#T_{j(b,i+1)-1}\right) + \#T_{j(a_{i+1},i+1)-1}.$$

The number of subwords in L_k of w is given by $1 + \#T_0$.

8. GENERALIZATIONS: THE FIBONACCI CASE

Other generalizations of the integer base numeration system are the linear numeration systems. In the previous sections, we considered the language $\{\varepsilon\} \cup 1\{0,1\}^*$. This means that there was no restriction on the words over $\{0,1\}$ except for the presence of a convenient leading 1. One could naturally define other Pascal triangles by considering any infinite language K over a totally ordered alphabet. If we order the words of K by genealogical ordering: $w_0, w_1, w_2, \ldots$, we obtain an infinite array where the element in the ith row and jth column, is $\binom{w_i}{w_j}$. This corresponds exactly to the case where we replace base-2 expansions with representations within an abstract numeration system based on K [23], [7, Chap. 3].

We will handle the case of the Fibonacci numeration system, i.e., with the language $L_F = \{\varepsilon\} \cup 1\{0,01\}^*$. It turns out that the sequence $(S_F(n))_{n \geq 0}$ counting the number of words in L_F occurring as subwords of the nth word in L_F has properties similar to those of $(S(n))_{n \geq 0}$. In particular, the notion of k-regularity may be extended to take into account a larger class of numeration systems [32, 1].

As observed in Remark 36, what seems to be important for further generalizations is that the language K of the numeration is the set of words not starting with 0 and not containing occurrences of a set of words of length 2. For the Fibonacci numeration system, the language of the numeration is obtained by avoiding the factor 11.

In the following definition, we recall the notion of a positional numeration system based on a sequence of integers. See, for instance, [16, 29] or [7, Chap. 2].

Definition 30. Let $U = (U(n))_{n \geq 0}$ be a sequence of integers such that U is increasing, $U(0) = 1$ and $\sup_{n \geq 0} \frac{U(n+1)}{U(n)}$ is bounded by a constant. We say that U is a *linear numeration system* if U satisfies a linear recurrence relation, i.e., there exist $a_0, \ldots, a_{k-1} \in \mathbb{Z}$ such that

$$\forall n \geq 0, \quad U(n+k) = a_{k-1} U(n+k-1) + \cdots + a_0 U(n).$$

Let n be a positive integer. By successive Euclidean divisions, there exists $\ell \geq 1$ such that

$$n = \sum_{j=0}^{\ell-1} c_j U(j)$$

where the c_j's are non-negative integers and $c_{\ell-1}$ is non-zero. The word $c_{\ell-1} \cdots c_0$ is called the *normal U-representation* of n and is denoted by $\text{rep}_U(n)$. In other words, the word $c_{\ell-1} \cdots c_0$ is the greedy expansion of n in the considered numeration. We set $\text{rep}_U(0) = \varepsilon$. Finally, we say that $\text{rep}_U(\mathbb{N})$ is the *language of the numeration.* If $d_{\ell-1} \cdots d_0$ is a word over an alphabet of digits, then we set

$$\text{val}_U(d_{\ell-1} \cdots d_0) = \sum_{j=0}^{\ell-1} d_j U(j).$$

We consider the Fibonacci sequence $F = (F(n))_{n \geq 0}$ defined by $F(0) = 1$, $F(1) = 2$ and $F(n+2) = F(n+1) + F(n)$ for all $n \geq 0$. Using the previous definition, the linear numeration system associated with F is called the *Zeckendorf numeration system* [34] or the *Fibonacci numeration system.* Observe that, for this system, $\text{rep}_F(\mathbb{N}) = L_F$ is given by the set of words over $\{0, 1\}$ avoiding the factor 11. In Table 2, we write the normal F-representations of the first few integers.

0	ε	6	1001	12	10101
1	1	7	1010	13	100000
2	10	8	10000	14	100001
3	100	9	10001	15	100010
4	101	10	10010	16	100100
5	1000	11	10100	17	100101

TABLE 2. The normal F-representations of the first few integers.

To define an array, we consider all the words in L_F and we order them using the radix order. In the same way as the generalized Pascal triangle P_2 defined in [24], for a word $u \in L_F$, we compute the binomial coefficient $\binom{u}{v}$ for all words $v \in L_F$. The first few values of the corresponding Pascal triangle with words in L_F, denoted by P_F, are given in Table 3.

As before, we are interested in a particular sequence.

Definition 31. Let $(S_F(n))_{n \geq 0}$ be the sequence whose nth term, $n \geq 0$, is the number of non-zero elements in the nth row of P_F. Hence, the first few terms of this sequence are

$$1, 2, 3, 4, 4, 5, 6, 6, 6, 8, 9, 8, 8, 7, 10, 12, 12, 12, 10, 12, 12, 8, 12, 15, 16, 16, 15, \ldots.$$

		ε	1	10	100	101	1000	1001	1010	10000	S_F
$\mathrm{rep}_F(0)$	ε	1	0	0	0	0	0	0	0	0	1
$\mathrm{rep}_F(1)$	1	1	1	0	0	0	0	0	0	0	2
$\mathrm{rep}_F(2)$	10	1	1	1	0	0	0	0	0	0	3
$\mathrm{rep}_F(3)$	100	1	1	2	1	0	0	0	0	0	4
$\mathrm{rep}_F(4)$	101	1	2	1	0	1	0	0	0	0	4
$\mathrm{rep}_F(5)$	1000	1	1	3	3	0	1	0	0	0	5
$\mathrm{rep}_F(6)$	1001	1	2	2	1	2	0	1	0	0	6
$\mathrm{rep}_F(7)$	1010	1	2	3	1	1	0	0	1	0	6
$\mathrm{rep}_F(8)$	10000	1	1	4	6	0	4	0	0	1	6

TABLE 3. The first few values in the generalized Pascal triangle P_F with words in L_F.

Otherwise stated, for $n \geq 0$,

$$(8) \qquad S_F(n) := \# \left\{ v \in L_F \mid \binom{\mathrm{rep}_F(n)}{v} > 0 \right\}.$$

In Figure 8, we have depicted the sequence $(S_F(n))_{n \geq 0}$ in the interval $[0, F(11)]$.

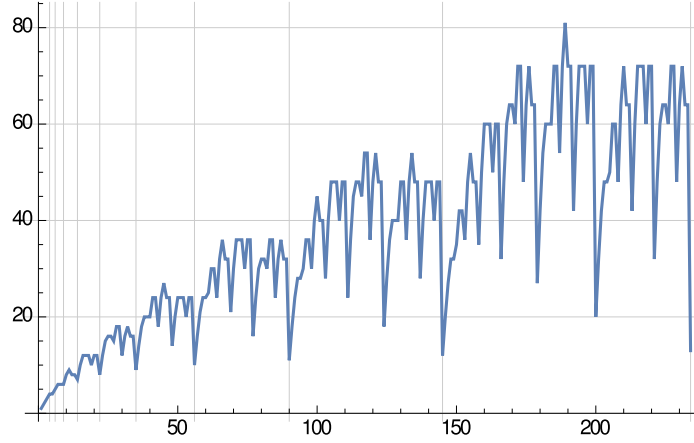

FIGURE 8. The sequence $(S_F(n))_{n \geq 0}$ in the interval $[0, F(11)]$.

Surprisingly, we will show that the sequence $(S_F(n))_{n \geq 0}$ satisfies a recurrence relation of the same kind as in Proposition 12 and is also regular in some extended way [1, 30].

Proposition 32. *We have $S_F(0) = 1$, $S_F(1) = 2$. Any integer $n \geq 2$, can be written as $F(\ell) + r$ for some $\ell \geq 1$ and $0 \leq r < F(\ell - 1)$. We have*

$$S_F(F(\ell) + r) = \begin{cases} S_F(F(\ell - 1) + r) + S_F(r), & \text{if } 0 \leq r < F(\ell - 2); \\ 2S_F(r), & \text{if } F(\ell - 2) \leq r < F(\ell - 1). \end{cases}$$

The proof is postponed to the end of the section (after Corollary 42). Using this result, there is a convenient way given in Table 4 to arrange the terms of the

1										
2										
3										
4	4									
5	6	6								
6	8	9	8	8						
7	10	12	12	12	10	12	12			
8	12	15	16	16	15	18	18	12	16	18	16	16

TABLE 4. Arrangement of the first few terms in $(S_F(n))_{n \geq 0}$.

sequence. The 0th (resp., first) row of this table contains the element $S_F(0)$ (resp., $S_F(1)$) of the sequence. Then, for all $n \geq 2$, the nth row contains the elements $S_F(i)$ where $i \in \{F(n-1), \ldots, F(n-1) + F(n-2) - 1\}$ and thus contains $F(n-2)$ elements. Using Proposition 32, for $n \geq 4$, the first $F(n-3)$ elements in the nth row are derived from the previous row: the difference of two consecutive rows is a prefix of $(S_F(n))_{n \geq 0}$. The last $F(n-4)$ elements in the nth row are twice the elements of the $(n-2)$th row. For $n \geq 1$, also observe that the first element of the nth row is equal to $S_F(F(n-1))$.

9. F-REGULARITY OF $(S_F(n))_{n \geq 0}$

We now study the regularity of the sequence $(S_F(n))_{n \geq 0}$. Observe that in Definition 19, to obtain one element of the kernel $\mathcal{K}_k(s)$, it is equivalent to consider a word $q \in \{0, \ldots, k-1\}^*$ and all the indices of s whose base-k expansion (with possibly some leading zeroes) ends with the suffix q. As an example, $(s(2^3 n + 1))_{n \geq 0}$ corresponds to the word $q = 001$ (in base 2).

Definition 33. Let X be a subset of $\mathbb{N}$. For each q in $\{0, 1\}^*$, we define the map

$$i_q(X) := \mathrm{val}_F(0^* \mathrm{rep}_F(X) \cap \{0, 1\}^* q).$$

In other words, i_q selects elements in X whose normal F-representation (padded with leading zeroes) ends with q. If $i_q(\mathbb{N})$ is non-empty, it is naturally ordered:

$$i_q(\mathbb{N}) = \{x_{q,0} < x_{q,1} < x_{q,2} < \cdots\}$$

and by abuse of notation, we set $i_q(n) = x_{q,n}$ for all $0 \leq n < \#i_q(\mathbb{N})$.

Example 34. The first values in $i_0(\mathbb{N})$ (resp., $i_1(\mathbb{N})$) are $0, 2, 3, 5, 7, 8$ (resp., $1, 4, 6, 9$). In particular, $i_0(0) = 0$, $i_0(1) = 2$, ..., $i_0(5) = 8$. The first values in $i_{10}(\mathbb{N})$ are $2, 7$.

The second part of the next lemma is particularly important when considering elements of the F-kernel (see Theorem 41).

Lemma 35. *We have $i_{pq}(\mathbb{N}) \subseteq i_q(\mathbb{N})$ and*

$$i_{pq}(\mathbb{N}) = i_{pq}(i_q(\mathbb{N})).$$

Moreover, if $pq \in L_F$, then

$$\mathrm{rep}_F(i_p(\mathbb{N}))q = \mathrm{rep}_F(i_{pq}(\mathbb{N}))$$

i.e., if $up \in 0^* L_F$ is such that $\mathrm{val}_F(up)$ is the nth value $x_{p,n}$ occurring in $i_p(\mathbb{N})$, then $\mathrm{val}_F(upq)$ is the nth value $x_{pq,n}$ occurring in $i_{pq}(\mathbb{N})$.

Let us illustrate the second part of Lemma 35 in Table 5. We can directly determine the nth word ending with 01 in the language of the numeration, from the nth word ending with 0 by simply adding a suffix 1.

$x_{0,0}$	ε	1	$x_{01,0}$
$x_{0,1}$	10	101	$x_{01,1}$
$x_{0,2}$	100	1001	$x_{01,2}$
$x_{0,3}$	1000	10001	$x_{01,3}$
$x_{0,4}$	1010	10101	$x_{01,4}$
$x_{0,5}$	10000	100001	$x_{01,5}$

TABLE 5. Illustration of Lemma 35.

Remark 36. In formal language theory, recall that Lw^{-1} denotes the set of words v such that $vw \in L$. The second part of Lemma 35 holds because the words in L_F are defined by avoiding the factor 11. Indeed, since 11 has length 2, we have $0^*L_F p^{-1} = 0^*L_F (pq)^{-1}$ when $pq \in L_F$ and $p \neq \varepsilon$.

This does not always hold, notably when there are longer forbidden factors in the language of the numeration. Even for the Tribonacci language L_T avoiding the factor 111, we have $1 \in 0^*L_T 1^{-1}$ and $1 \notin 0^*L_T (11)^{-1}$.

Definition 20 is replaced for the Fibonacci numeration system by the following notion where subsequences of a given sequence are selected by suffix of normal F-representations. This extension was first introduced in [32, 1].

Definition 37. Let q be a word in $\{0,1\}^*$ such that $i_q(\mathbb{N}) \neq \emptyset$ and let $s = (s(n))_{n \geq 0}$ be a sequence. The subsequence of s defined by $n \mapsto s(i_q(n))$ is called the *subsequence* of s *with least significant digits equal to q*. The set of all these subsequences when q belongs to $\{0,1\}^*$ and is such that $i_q(\mathbb{N}) \neq \emptyset$ is called the *Fibonacci-kernel* or F-*kernel* of the sequence s and is denoted by $\mathcal{K}_F(s)$. We say that s is *Fibonacci-regular* or F-*regular* if $\langle \mathcal{K}_F(s) \rangle$ is a finitely-generated $\mathbb{Z}$-module.

Each non-empty word in L_F is factorized into consecutive maximal blocks of zeroes separated by a unique one. It is of the form

$$10^{n_k} 10^{n_{k-1}} \cdots 10^{n_2} 10^{n_1}$$

with $n_1 \geq 0$ and $n_2, \ldots, n_k > 0$. Depending on the form of $u \in L_F$, we have a formula to count the number of words $v \in L_F$ such that $\binom{u}{v} > 0$.

Proposition 38. *Let u be a non-empty word in L_F of the form $10^{n_k} 10^{n_{k-1}} \cdots 10^{n_2} 10^{n_1}$ with $n_1 \geq 0$ and $n_2, \ldots, n_k > 0$. Then*

$$\#\left\{ v \in L_F \mid \binom{u}{v} > 0 \right\} = (n_1 + 2) \cdot \prod_{j=2}^{k} (n_j + 1).$$

To prove this result, we essentially follow the same ideas as for Corollary 9 but we need to consider the tree of subwords restricted to L_F.

For a word $w \in L_F$, the tree $\mathcal{T}_{L_F}(w)$ is given in Definition 3. The factorization (3) of words in L_F has a very particular form (because there is no 11). To refer to

the same subtrees as in Definition 6, we stick to the notation of (3) even though
the blocks of ones are limited to a single digit

$$w = \underbrace{1}_{u_1}\underbrace{0^{n_2}}_{u_2}\underbrace{1}_{u_3}\underbrace{0^{n_4}}_{u_4}\cdots\underbrace{1}_{u_{2j-1}}\underbrace{0^{n_{2j}}}_{u_{2j}}$$

with $j \geq 1$, $n_2, \ldots, n_{2j-2} \geq 1$ and $n_{2j} \geq 0$. Let $M = M_w$ be such that $w = u_1 u_2 \cdots u_M$ where u_M is the last non-empty block of zeroes or the last one.

Example 39. Consider the word $w = 101000100$. With the above notation, $M = 6$. In Figure 9, we have represented the trie of subwords $\mathcal{T}_{L_F}(w)$ in L_F and the subtrees $T_0, \ldots, T_5$. The roots of these subtrees correspond to a prefix of w ending with 1 or 10. Figure 10 illustrates the next proposition. We see how the subtrees are
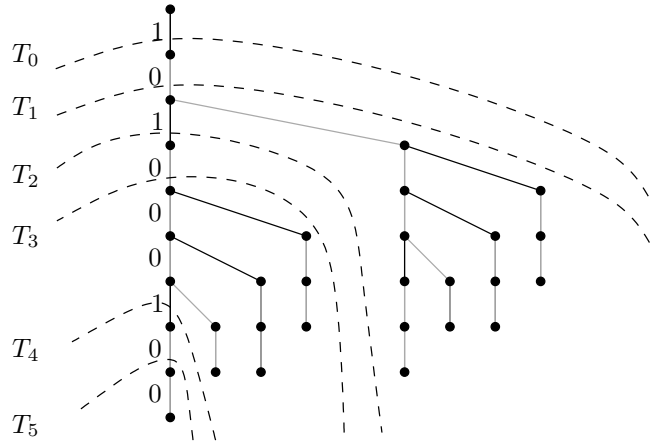
FIGURE 9. The trie $\mathcal{T}_{L_F}(101000100)$.

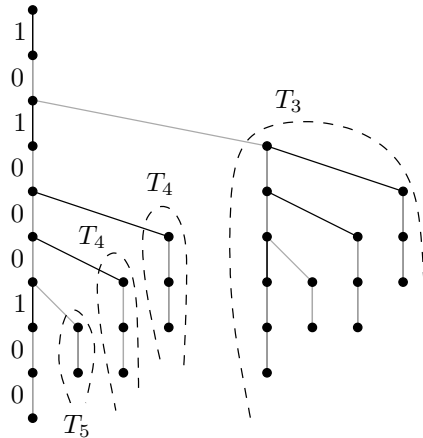connected to the "initial" linear subtree labeled by w.

FIGURE 10. The trie $\mathcal{T}_{L_F}(101000100)$.

Since we are considering the language L_F, the analogue of Proposition 7 becomes:

Proposition 40. *Let w be a finite word in L_F. With the above notation about M and the subtrees T_ℓ, the tree $\mathcal{T}_{L_F}(w)$ has the following properties.*

- *Assume that $2 \le 2k < M$. For every $j \in \{1, \ldots, n_{2k} - 1\}$, the node of label $x = u_1 \cdots u_{2k-1} 0^j$ has two children $x0$ and $x1$. The node $x1$ is the root of a tree isomorphic to T_{2k}. Moreover, $x = u_1 \cdots u_{2k-1}$ has a single child $x0$.*
- *Assume that $3 \le 2k + 1 < M$. The node of label $x = u_1 \cdots u_{2k}$ has two children $x0$ and $x1$. The node $x0$ is the root of a tree isomorphic to T_{2k+1}.*
- *The node of label $x = \varepsilon$ has only one child $x1$.*

Proof of Proposition 38. Proceed by induction and consider two cases depending on the last letter of u. Start with a linear tree and add, with a bottom-up approach, all the possible subtrees given by Proposition 38. $\qquad\square$

Theorem 41. *The sequence $(S_F(n))_{n \ge 0}$ satisfies, for all $n \ge 0$,*

$$S_F(i_{00}(n)) = 2S_F(i_0(n)) - S_F(i_\varepsilon(n))$$
$$S_F(i_{01}(n)) = 2S_F(i_\varepsilon(n))$$
$$S_F(i_{010}(n)) = 3S_F(i_\varepsilon(n)).$$

In particular, $(S_F(n))_{n \ge 0}$ is F-regular.

Proof. Let $q \in \{0, 1\}^*$ such that $i_q(\mathbb{N}) \ne \emptyset$. Recall that, for all $n \ge 0$,

$$S_F(i_q(n)) = \#\left\{v \in L_F \mid \binom{\mathrm{rep}_F(i_q(n))}{v} > 0\right\}.$$

Let us prove the first relation. Let u be a non-empty word in L_F. The idea is to use Proposition 38. Let us write u as $10^{n_k} 10^{n_{k-1}} \cdots 10^{n_1}$ with $n_2, \ldots, n_k > 0$ and $n_1 \ge 0$. By Proposition 38 we have

$$\#\left\{v \in L_F \mid \binom{u}{v} > 0\right\} = (n_1 + 2) \cdot \prod_{j=2}^{k} (n_j + 1)$$

$$\#\left\{v \in L_F \mid \binom{u0}{v} > 0\right\} = (n_1 + 3) \cdot \prod_{j=1}^{k} (n_j + 1)$$

$$\#\left\{v \in L_F \mid \binom{u00}{v} > 0\right\} = (n_1 + 4) \cdot \prod_{j=1}^{k} (n_j + 1).$$

Hence, we get

$$2 \cdot \#\left\{v \in L_F \mid \binom{u0}{v} > 0\right\} - \#\left\{v \in L_F \mid \binom{u}{v} > 0\right\} = \#\left\{v \in L_F \mid \binom{u00}{v} > 0\right\}.$$

This leads to the expected relation because

$$\mathrm{val}_F(\mathrm{rep}_F(n)0) = i_0(n) \text{ and } \mathrm{val}_F(\mathrm{rep}_F(n)00) = i_{00}(n).$$

These latter relations are reflected by the second part of Lemma 35.

The last two relations are obtained using the same ideas. One has to simply use Proposition 38 with words of the form u, $u01$ and $u010$. We derive that

$$\#\left\{v \in L_F \mid \binom{u01}{v} > 0\right\} = 2 \cdot \#\left\{v \in L_F \mid \binom{u}{v} > 0\right\}$$

and

$$\# \left\{ v \in L_F \mid \binom{u010}{v} > 0 \right\} = 3 \cdot \# \left\{ v \in L_F \mid \binom{u}{v} > 0 \right\}.$$

The F-regularity of the sequence follows from Lemma 35: the $\mathbb{Z}$-module generated by the F-kernel of S_F is generated by $(S_F(i_\varepsilon(n)))_{n \geq 0} = (S_F(n))_{n \geq 0}$ and $(S_F(i_0(n)))_{n \geq 0}$. As an example,

$$(S_F(i_{1001}(n)))_{n \geq 0} = 2(S_F(i_{10}(n)))_{n \geq 0} = 2(S_F(i_{010}(n)))_{n \geq 0} = 6(S_F(n))_{n \geq 0}.$$

$\square$

Corollary 22 is replaced by the following result, which says that if a sequence is F-regular, then its nth term can be obtained by multiplying some matrices. The length of this product is proportional to $\log_\varphi(n)$ where φ is the golden ratio. Here again, we get matrices of size 2 thanks to Theorem 41.

Corollary 42. *Let*

(9) $$V(0) = \begin{pmatrix} S_F(i_\varepsilon(0)) \\ S_F(i_0(0)) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Consider the matrix-valued morphism $\mu : \{0, 01\}^* \to \mathbb{Z}_2^2$ *defined by*

$$\mu(0) = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}, \quad \mu(01) = \begin{pmatrix} 2 & 0 \\ 3 & 0 \end{pmatrix}.$$

For all $n \geq 0$, let $0\,\mathrm{rep}_F(n) = u_k \cdots u_1$ where $u_i \in \{0, 01\}$ for all $i \in \{1, \ldots, k\}$. Then

$$S_F(n) = \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(u_1) \cdots \mu(u_k) \, V(0).$$

Proof. We proceed by induction on k in the factorization $0\,\mathrm{rep}_F(n) = u_k \cdots u_1$ with $u_i \in \{0, 01\}$ for all i. One can observe that the result is true for $k \in \{1, 2\}$.

Assume now that $0\,\mathrm{rep}_F(n) = u_k \cdots u_1$ with $k \geq 3$. We only consider the case $u_2 = 0$, the other one is similar. If $u_1 = 0$, then $S_F(n) = S_F(i_{00}(m))$ for some $m \geq 0$. By Theorem 41 and using the induction hypothesis, we get

$$\begin{aligned} S_F(n) &= 2S_F(i_0(m)) - S_F(i_\varepsilon(m)) \\ &= 2S_F(\mathrm{val}_F(u_k \cdots u_2)) - S_F(\mathrm{val}_F(u_k \cdots u_3)) \\ &= \begin{pmatrix} 1 & 0 \end{pmatrix} (2\mu(0) - I) \, \mu(u_3) \cdots \mu(u_k) \, V(0). \end{aligned}$$

The equality follows by observing that $(2\mu(0) - I) = \mu(0)^2$. If $u_1 = 01$, then $S_F(n) = S_F(i_{01}(m))$ for some $m \geq 0$. By Theorem 41 and using the induction hypothesis, we get

$$\begin{aligned} S_F(n) &= 2S_F(i_\varepsilon(m)) \\ &= 2S_F(\mathrm{val}_F(u_k \cdots u_2)) \\ &= 2 \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(u_2) \cdots \mu(u_k) \, V(0). \end{aligned}$$

The equality follows by observing that $2 \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(01)$. $\square$

Proof of Proposition 32. We make use of the previous corollary. Assume that $n = F(\ell) + r$ with $\ell \geq 1$ and $0 \leq r < F(\ell - 1)$. We have $0\,\mathrm{rep}_F(n) = u_k \cdots u_1$ for some $k \geq 2$, with $u_i \in \{0, 01\}$ for all i.

If $F(\ell - 2) \le r < F(\ell - 1)$, then $u_{k-1} = 01$. By Corollary 42, we get

$$
\begin{aligned}
S_F(F(\ell) + r) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(u_1) \cdots \mu(u_{k-2})\mu(01)\mu(01)\, V(0); \\
S_F(r) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(u_1) \cdots \mu(u_{k-2})\mu(01)\, V(0)
\end{aligned}
$$

and the equality $S_F(F(\ell) + r) = 2S_F(r)$ follows.

If $0 \le r < F(\ell - 2)$, then $u_{k-1} = 0$. Let $m < k - 1$ be the greatest integer such that $u_m = 01$ (we set $m = 0$ if $u_i = 0$ for all $i \le k - 1$). By Corollary 42, we have

$$
\begin{aligned}
S_F(F(\ell) + r) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(u_1) \cdots \mu(u_m)\mu(0)^{k-m-1}\mu(01)\, V(0); \\
S_F(F(\ell-1) + r) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(u_1) \cdots \mu(u_m)\mu(0)^{k-m-2}\mu(01)\, V(0); \\
S_F(r) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \mu(u_1) \cdots \mu(u_m)\, V(0).
\end{aligned}
$$

We get the equality $S_F(F(\ell)+r) = S_F(F(\ell-1)+r)+S_F(r)$ by showing by induction that, for all n,

$$
\mu(0)^n \mu(01) = \begin{pmatrix} n + 2 & 0 \\ n + 3 & 0 \end{pmatrix}.
$$

$\square$

10. Concluding remarks

After considering the language L_2 of base-2 expansions, then the language L_F of Fibonacci expansions, one can naturally wonder whether similar properties can be observed for an arbitrary initial language (because Pascal-like triangles may be defined in this general setting). A first generalization of the Fibonacci case would be to consider the m-bonacci case where the corresponding language is made of the words over $\{0, 1\}$ avoiding the factor 1^m (the Fibonacci case is $m = 2$ and the Tribonacci case $m = 3$ has been considered in Remark 36).

Example 43. For $m = 3$, the sequence counting admissible subwords associated with the Tribonacci numeration system starts with

$$1, 2, 3, 3, 4, 5, 5, 5, 7, 8, 6, 7, 7, 6, 9, 11, 9, 11, 12, 10, 9, 11, 11, 9, 7, 11, 14, 12, 15,$$

$$17, 15, 14, 18, 19, 15, 14, 14, 11, 15, 17, 15, 15, 17, 15, 8, 13, 17, 15, 19, 22, \ldots.$$

Due to Remark 36 that extends easily to m-bonacci languages for $m > 3$, Lemma 35 does not hold for the m-bonacci numeration system as soon as $m \ge 3$. Consequently, it is not clear whether the sequence $(S_T(n))_{n \ge 0}$ is T-regular or the analogue sequence for the m-bonacci case is m-bonacci-regular. Nevertheless, the sequence S_T seems to partially satisfy a relation similar to the first part of Proposition 32. To build a table similar to the arrangement found in Table 4, numerical observations lead to the following conjecture: if n_i denotes the position of the last occurrence of i in S_T (assuming that n_i is thus well defined, which is the case for the Fibonacci language, as shown by Proposition 44 below), then $S_T(n_i + r) = S_T(n_{i-1} + r) + S_T(r)$ for $0 \le r < n_i - n_{i-1}$, $i \ge 5$. Moreover, the sequence

$$(n_i)_{i \ge 1} = 0, 1, 3, 4, 7, 13, 24, 44, 81, 149, 274, 504, \ldots$$

satisfies the same linear relation as the Tribonacci sequence when $i \ge 4$. Nevertheless, it is not clear that one can determine a "simple" relation for $S_T(n_i + r)$ when $n_i - n_{i-1} \le r < n_{i+1} - n_i$ (corresponding to the second part of Proposition 32) and thus derive a possible regularity of the sequence S_T. Again, one can also try with larger values of the parameter m and imagine partial relations of the same form.

```
1
2   3
3
4   5   5
5   7   8   6   7   7
6   9   11  9   11  12  10  9   11  11  9
7   11  14  12  15  17  15  14  18  19  15  14  14  11  15  17  15  15  17  15
8   13  17  15  19  22  ...
```

TABLE 6. Arrangement of the first few terms in $(S_T(n))_{n\geq 0}$.

In the Fibonacci case, if n_i denotes the position of the last occurrence of i in $(S_F(n))_{n\geq 0}$ for all $i \geq 1$, then $S_F(n_i+r) = S_F(n_{i-1}+r)+S_F(r)$ for $0 \leq r < n_i-n_{i-1}$ and $i \geq 5$. Indeed, we just need to combine Proposition 32 and Proposition 44.

Proposition 44. *For all $i \geq 1$, the number of occurrences of i in $(S_F(n))_{n\geq 0}$ is finite. If n_i denotes the position of the last occurrence of i in $(S_F(n))_{n\geq 0}$, then $n_i = F(i-2)$ for all $i \geq 5$.*

Proof. First of all, Proposition 32 implies that

$$(10) \qquad\qquad S_F(n) > 1 = S_F(0) \quad \forall n \geq 1.$$

Then for all $\ell \geq 0$, using Proposition 32 leads us to

$$(11)\quad S_F(F(\ell)) = S_F(F(\ell-1))+1 = S_F(F(\ell-2))+2 = \cdots = S_F(F(0))+\ell = 2+\ell.$$

Now we prove the following result : for all $\ell \geq 3$, $S_F(F(\ell))$ is less than $S_F(F(\ell)+r)$ for $0 < r < F(\ell-1)$. We show this by induction on ℓ. If $\ell \in \{3,4\}$, then Table 4 gives the result. We suppose the result holds true up to $\ell-1$ and we show it also holds true for ℓ. For $0 < r < F(\ell-2)$, we have

$$
\begin{aligned}
S_F(F(\ell)+r) \;&=\; S_F(F(\ell-1)+r)+S_F(r) &&\text{(by Proposition 32)}\\
&>\; S_F(F(\ell-1))+S_F(0) &&\text{(by induction hypothesis and by (10))}\\
&>\; S_F(F(\ell)) &&\text{(by Proposition 32)}.
\end{aligned}
$$

Now, for $F(\ell-2) \leq r < F(\ell-1)$, we have $S_F(F(\ell)+r) = 2S_F(r)$ by Proposition 32. There exists $0 \leq r' < F(\ell-3)$ such that $r = F(\ell-2)+r'$. We get

$$
\begin{aligned}
S_F(F(\ell)+r) \;&\geq\; 2S_F(F(\ell-2)) &&\text{(by induction hypothesis)}\\
&\geq\; 2\ell &&\text{(by (11))}\\
&>\; 2+\ell &&\text{(since $\ell \geq 5$)}\\
&>\; S_F(F(\ell)) &&\text{(by (11))}.
\end{aligned}
$$

This ends the proof of the intermediate result. Observe that, if $\ell \geq 3$, we have that $S_F(F(\ell)) < S_F(n)$ for all $n > F(\ell)$ since $S_F(F(\ell)) < S_F(F(\ell+m))$ with $m \geq 1$.

For all $i \geq 1$, we conclude that the number of occurrences of i in $(S_F(n))_{n\geq 0}$ is finite. The sequence $(n_i)_{i\geq 1}$ is thus well-defined. We also get that $n_i = F(i-2)$ for all $i \geq 5$. □

References

[1] J.-P. Allouche, K. Scheicher, R. F. Tichy, Regular maps in generalized number systems, *Math. Slovaca* **50** (2000), 41–58.

[2] J.-P. Allouche, J. Shallit, The ring of k-regular sequences, *Theoret. Comput. Sci.*, **98** (1992), 163–197.

[3] J.-P. Allouche, J. Shallit, *Automatic sequences. Theory, applications, generalizations*, Cambridge University Press, (2003).

[4] B. Bates, M. Bunder, K. Tognetti, Locating terms in the Stern–Brocot tree, *European J. Combin.* **31** (2010), 1020–1033.

[5] J. Berstel, An Exercise on Fibonacci Representations, *Theoret. Inform. Appl.* **35** (2001), no. 6, 491–498.

[6] J. Berstel, C. Reutenauer, *Noncommutative rational series with applications*, Encycl. of Math. and its Appl. **137**, Cambridge Univ. Press, (2011).

[7] V. Berthé, M. Rigo (Eds.), *Combinatorics, automata and number theory*, Encycl. of Math. and its Appl. **135**, Cambridge Univ. Press, (2010).

[8] C. Bright, R. Devillers, J. Shallit, Subwords, regular languages, and prime numbers, *Exp. Math.* **25** (2016), no. 3, 321–331.

[9] N.J. Calkin, H.S. Wilf, Binary partitions of integers and Stern–Brocot-like trees (1998), unpublished. Updated version August 5, 2009, 19 pages.

[10] A. Carpi, C. Maggi, On synchronized sequences and their separators, *Theor. Inform. Appl.* **35** (2001), 513–524.

[11] M. Coons, J. Shallit, A pattern sequence approach to Stern's sequence, *Discrete Math.* **311** (2011), no. 22, 2630–2633.

[12] P. Dumas, Joint spectral radius, dilation equations, and asymptotic behavior of radix-rational sequences, *Linear Algebra Appl.* **438** (2013), no. 5, 21072126.

[13] A. W. M. Dress, P. L. Erdős, Reconstructing words from subwords in linear time, *Annals of Combinatorics* **8** (2004), 457–462.

[14] C. H. Du, H. Mousavi, L. Schaeffer, J. Shallit, Decision Algorithms for Fibonacci-Automatic Words, III: Enumeration and Abelian Properties, preprint (2016).

[15] S. Eilenberg, *Automata, Languages and Machines*, Vol. A, Academic Press, New York (1974).

[16] A. S. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985), 105–114.

[17] D. D. Freydenberger, P. Gawrychowski, J. Karhumäki, F. Manea, W. Rytter, Testing k-binomial equivalence, `arXiv:1509.00622`.

[18] S. P. Glasby, Enumerating the rationals from left to right, *Amer. Math. Monthly* **118** (2011), 830–835.

[19] D. Goč, L. Schaeffer, J. Shallit, Subword complexity and k-synchronization, in Developments in language theory, *Lect. Notes in Comput. Sci.* **7907** (2013), 252–263.

[20] F. Greinecker, On the 2-abelian complexity of the Thue-Morse word, *Theoret. Comput. Sci.* **593** (2015), 88–105.

[21] P. Karandikar, M. Kufleitner, Ph. Schnoebelen, On the index of Simon's congruence for piecewise testability, *Inform. Processing Let.* **15** (2015), 515–519.

[22] P. Karandikar, M. Niewerth, Ph. Schnoebelen, On the state complexity of closures and interiors of regular languages with subwords and superwords, *Theoret. Comput. Sci.* **610** (2016), 91–107.

[23] P. B. A. Lecomte, M. Rigo, Numeration systems on a regular language, *Theory Comput. Syst.* **34** (2001), 27–44.

[24] J. Leroy, M. Rigo, M. Stipulanti, Generalized Pascal triangle for binomial coefficients of words, *Adv. in Appl. Math.* **80** (2016), 24–47.

[25] J. Leroy, M. Rigo, M. Stipulanti, Summatory function of sequences counting subwords occurrences (2016), submitted preprint.

[26] M. Lothaire, *Combinatorics on Words*, Cambridge Mathematical Library, Cambridge University Press, (1997).

[27] A. Parreau, M. Rigo, E. Rowland, É. Vandomme, A new approach to the 2-regularity of the ℓ-abelian complexity of 2-automatic sequences, *Electron. J. Combin.* **22** (2015), 1–27.

[28] M. Rao, M. Rigo, P. Salimov, Avoiding 2-binomial squares and cubes, *Theoret. Comput. Sci.* **572** (2015), 83–91.

[29] M. Rigo, *Formal languages, automata and numeration systems, vol. 1, Introduction to combinatorics on words*, ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, (2014).

[30] M. Rigo, A. Maes, More on generalized automatic sequences, *J. Autom. Lang. Comb.* **7** (2002), 351–376.

[31] L. Schaeffer, J. Shallit, Closed, Palindromic, Rich, Privileged, Trapezoidal, and Balanced Words in Automatic Sequences, *Electron. J. Comb.* **23** (2016), no. 1, paper 1.25.

[32] J. Shallit, A generalization of automatic sequences, *Theoret. Comp. Sci.* **61** (1988), no. 1, 1–16.

[33] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, `oeis.org`.

[34] É. Zeckendorf, Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas, *Bull. Soc. Roy. Sci. Liège* **41** (1972), 179–182.

Université de Liège, Institut de mathématique, Allée de la découverte 12 (B37), 4000 Liège, Belgium

J.Leroy@ulg.ac.be, M.Rigo@ulg.ac.be, M.Stipulanti@ulg.ac.be